# BEHAVIOURAL DECISIONS: THEORY, IMPLICATIONS AND APPLICATIONS

**by**

**Patricio Santiago Dalton**

A Thesis submitted in partial fulfilment of the requirements
for the degree of Doctor of Philosophy in Economics

University of Warwick
Department of Economics
July 2009

# Contents

# List of Figures

# List of Tables

# Acknowledgments

> "... y desafiando el oleaje sin timón ni timonel, por mis sueños va, ligero de equipaje, sobre un cascaron de nuez, mi corazón de viaje, luciendo los tatuajes de un pasado bucanero..." (Peces de ciudad, Joaquín Sabina)

"There are those who know from the start where they are going and those who only realize after the journey where they have been travelling" Dworkin (1988, pg. ix). I contain a bit of both types. I knew my destination from the beginning of this Ph.D. journey: I wanted to bring a pinch of Psychology into Economic Theory. I did not imagine, however, what my journey would be like or whether I would get to the destination. I just liked the challenge and got on board, "desafiando el oleaje."

I did not depart alone from the harbour that is Buenos Aires. Malena was there with me. She fed my heart and enlightened my soul. She gave me the proof of the most difficult theorem. She proved that love has not boundaries. Her capacity to love from her deepest entrails was my main shelter for inspiration. Gracias amor, I love and admire you so much!

My family was also there in the boat with me, day and night. Barbi, with her music and personal example has been a very important source of inspiration. Her Bach's Prelude was always with me, and my research has been inspired by her work with "las orquestas." Lygia, with her wisdom and spirituality gave me the peace of mind needed in the difficult moments. Tati and Nani, you gave me all that I am: this thesis is also yours (the errors are mine). Ualdito, thanks for being always there. Abuela, no wonder why Male has such a big heart! Your prayer got always to me.

I brought my friends from life in the boat. Cristian Mazza (Titi), Juan Jose Cerutti (Juanjismo), Ariel Buongiorno (el Gordito), Mariano Gomez (Cabeza), Jose Noguera (Jose) fed the "barrio" part of my deepest self. A great deal of what I am and what I did here is due to you guys. Cachito Buongiorno was also there with me. Now he must be somewhere up in a cloud proud of his "Mingo." I also owe a great deal of this thesis to Patricia Vazquez, the main architect of my character. Thank you Mari for introducing me to Patri!

When a strong wind hit the boat, I was lucky enough to be with all my crew. Especially I had the fortune to have Isabella Sullis and Male with me at home. Isa, la "sorella," took great care of the "patient" with much love. Male showed me the complete proof of her theorem.

If I could see any further during this journey, is because I could stood on the shoulder of some giants who were willing to provide encouragement, guidance and wisdom. Top in that ranking is Sayantan Ghosal, who effectively provided me with these three essential inputs. Thank you Sayantan for trusting me, my capacity and my ideas from the very beginning. I still remember the day I told you that I wanted to apply for a Ph.D. in UK and you said to me "I am pleased to write a recommendation letter for you, but do consider seriously coming back to Warwick." You not only wrote the reference letter, but you put forward my name for the Marie Curie Scholarship. Thank you also for being so open minded and for believing in what I wanted to do. I owe you much of what I have learned during my research.

Saul Keifman and Daniel Heymann, my mentors in Argentina, were other giants on whose shoulders I stood up. I will always be grateful that you started me off on the path that led to this thesis and for your continuing support.

During this trip I got to know very nice and smart people who gave me feedback, guided me (I didn't have a GPS!) and, on occasions, offered me a life-jacket. I was extremely lucky to meet Peter Hammond, an adorable person. Since the very first moment I met him, he showed me how special he is. Thank you Peter for being such a passionate and open minded academic and for having been such a central support in my career! Thank you also for introducing me to Douglas Bernheim, who became a significant person at the end of this journey. I am indebted to Doug who "adopted" me as his "distant student" and supported me and my research extraordinarily.

Thanks are due to Andres (Abbondanzieri) Carvajal, a friend and another giant! I will never forget the help he gave me just when I needed it the most. Andres, your goodness of heart is definitively testable! Thank you so much. Anandi Mani and Norman Ireland have placed their trust in me and my project, writing up several recommendation letters. Andrew Oswald's and Mark Stwart's feedback on the empirical part of this thesis has been also very valuable. I am equally thankful to Carlo Perroni, who provided me with support during the initial stages of the research.

Being so far away from home would not have been possible without the invaluable capital of my friends in UK. Here I built up my Colombian family: Natalia Gonzalez (Nata) and Juan Gomez (Juanso), grandma and grandpa and more, your unbounded heart was essential in this trip. Mariana Blanco (Marian) and Juan Vargas (Juancin), my sister and brother and Andres and Ana Carvajal, my uncle and aunt. Thanks for adopting me as part of your family and for your warm friendship in the cold English winters. Nothing more comforting that our "comilonas" on weekends and your hugs! The branch of my family in Lyon was also extremely needed. Loezig Sevellec (Lou), Mariana Rimoldi (Mari) and Mael Sevellec (Mamel), your closeness is my treasure.

Chris Okeke (crizo), my Nigerian buddy; Gianluca Grimalda (GG), my most bizarre and lovely friend; Santiago Acosta (Santi), my Ph.D. co-fighter; Marcelo Saguier (Marce), my reflective buddy; Karen Van Rompaey (K), a piece of my "paisito" in Midlands; Andrea Salvatori (Salvatuts) and Flavia Moreschini (Fla), my great office and home friends; Luciana Juvenal (Lucci), my economic advisor; Sheheryar Malik (Sherry), my San Francisco's buddy; Mike Nibbet (Mike), the best native proof-reader; Manu & Jime, Miguel Ferreira, Carlos Ramirez, Federico Mateos and Analia Ruggiero. In different ways your friendship was a shelter to me during this journey, and sometimes was literally a roof under which to sleep. Thank you so much!

Finally, this beautiful journey couldn't have been made without the financial support of the EST Marie Curie Fellowship, The Royal Economic Society Junior Fellowship, Warwick Economics Department and the ESRC-DFID grant No. RES-167-25-0364. I am in debt to these institutions.

In the end, I am here in a safe and sunny harbour. It was a long way to go but it was well worth the trip. A dream has been achieved. Thanks are due to all of you!

# Dedication

This thesis is dedicated to my two loves: Malena Digiuni and Club Atletico San Lorenzo de Almagro.

# Declaration

I declare that any material contained in this thesis has not been submitted for a degree to any other university. I further declare that one paper titled: "Behavioural Decisions and Welfare", drawn from Chapter 2 and 3 of this thesis appears as a Working paper (WP) in The Warwick Economics Research Paper Series, TWERPS 834, 2008, co-authored with Prof. Sayantan Ghosal.

Patricio Santiago Dalton

July 2009

# Abstract

Standard Economics assumes that decision makers carefully reflect and meditate on each alternative. Grounded on evidence from Psychology this thesis relaxes this assumption and develops a theoretical framework with imperfect reflection. We analyse the implications for both Positive and Normative Economics, and for Public Policy and introduce novel applications. We begin in Chapter 1 with a review of the main arguments and results.

Chapter 2 develops a decision-making framework with imperfect reflection and studies the implications for Positive Economics. Our approach relies on two key premises: (a) preferences parameters are endogenously determined by choices and (b) the decision maker, due to imperfect reflection, may not fully internalize that effect. A decision problem is labelled "Standard" (SDP) when the decision maker internalizes the effect and "Behavioural" (BDP) when she does not. A number of general results are obtained. First, under incomplete and acyclic preferences there exists a solution to both SDP and BDP. Second, BDPs have testable restrictions and they are different from those of SDPs. Third, for almost all classes of preferences, SDP and BDP are distinguishable from each other.

Chapter 3 focuses on the normative implications of our framework in relation to the existing normative approaches to Behavioural Economics. We show that, with imperfect reflection, revealed preferences cannot in general underpin individual welfare, though we offer conditions to recover such a link. The degree of "autonomy" emerges as a natural normative criterion and offers theoretical grounds for public policy interventions that aim to empower individuals.

Chapter 4 extends our framework to an N-person strategic setting. We label "psycho-social" games to normal-form games in which players, due to imperfect reflection, do not internalize the effect of their actions on their preferences parameters. We prove existence under weak assumptions, link our framework to the existent literature on social preferences and provide new examples. We associate psycho-social games to standard normal-form games and show that, typically, the set of Nash equilibria and the set of psycho-social equilibria of an associated psycho-social game are distinct from each other. Finally, we show that (strict) Nash equilibria in pure strategies are robust to any degree of imperfect reflection.

Chapter 5 applies our framework to analyse poverty and polarization persistence as a result of an aspirations failure. Reflection involves imagination which is, in turn, needed to form aspirations: i.e. we can only aspire to what we can imagine. Initial circumstances such us poverty and polarization restrict the scope of people's imagination and consequently of their aspirations. We develop a model consistent with this argument and study the importance of role models to break aspirations failures.

Chapter 6 combines experimental data with a survey and shows evidence that the more frustrated people are income-wise, the lower is their propensity to cooperate to provide global (e.g. environment) and national public goods. Finally, chapter 7 concludes and highlights avenues for future research.

# Chapter 1

# Introduction

*"Most powerful is he who has himself in his own power"*

(Seneca, 5 BC - 65 AD)

The capacity to reflect is what distinguishes us from the rest of animals. As humans, we have the ability to consciously and intentionally transform the world. The rest of animals, in contrast, can alter aspects of the world in the process of adapting to it, but their modification of reality is purely instinctive (Roberts, 2000, pg. 20).

Standard Economics takes careful consideration of this capacity and provides the "homo-economicus" with the gift of perfect reflection. The decision maker is supposed to embark upon a conscious mental process to carefully characterize each alternative involved and fully evaluate all their consequences. If needed, the decision maker is supposed to actually imagine herself experiencing each of the alternatives. As pointed out by "the Bible" of any economics Ph.D. student, "our decision makers make only meditated choices" (Mas-Colell et al., 1995, pg. 6).

The strength of this assumption depends crucially on the complexity of the decision problem in question. When it comes to simple decision problems, perfect reflection would seem to be an innocuous assumption to make. In complex decisions, however, the level of reflection needed to characterize the object of choice is higher than in simple decisions, and the assumption of perfect reflection becomes difficult to justify.

At this point it is important to clarify what we mean by complexity of a decision. We define a complex decision problem as one in which (generally unobserved) payoff-relevant preference parameters are defined endogenously as part of the choice. Examples of such endogenous preferences parameters are mental states, reference points, frames, moods, beliefs, aspirations, attitudes, emotions, values or the environment itself. So what characterizes a complex decision is the presence of preference parameters that are essentially part of the consequences of the choice.

Many if not most real-life decisions are complex. Evidence from Psychology and Behavioural Economics shows that even those decisions that look trivial at first glance can be tricky. Think about the choice between an apple and a banana in your local grocery. This is the simplest decision we can think of, and as such, it requires almost no reflection: my pleasure when I eat an apple is greater than when I eat a banana, and therefore, I choose the apple when both are available. In such a simplistic world, the assumption of perfect reflection does not affect my definition of the objects of choice: i.e. banana and apple. Suppose that the greengrocer has read Sunstein and Thaler's (2008) book, *Nudge*, and discovers that the position in which the types of fruit are presented on the shelves affect the consumption of the bananas and apples. The decision problem becomes complex now, and the decision maker will require a higher level of reflection to fully define the object of choice, which is not a particular piece of fruit anymore, but a piece of fruit together with its position in the shelf. A decision maker with perfect reflection will, hopefully, understand this complex problem and maximize her preferences over pairs of "fruit and position on the shelf" and make consistent decision choices. If upon reflection, she ranks apples over bananas even when bananas are presented as default options, then her choices will not be liable to be manipulated by the greengrocer and she will consistently choose the apple. Our framework differs from Sunstein and Thaler's (2008) approach mainly in the fact that they propose a social planner as an "architect of the decision-maker choice." We propose, instead, that the decision maker herself should be the "architect of her own choice" or her "own character planner" in the words of Elster (1983). Our claim is not grounded only on "moral" considerations but also on practical concerns.

Think for a moment on what Sunstein and Thaler's "benevolent social planner" is supposed to do. He should (a) recognize the imperfect reflection of the decision maker, (b) identify that her true preferences favour apples over bananas and (c) present the different types of fruit in such a way that the decision maker chooses his apple. I ask to myself, wouldn't it be too demanding for a social planner to do so? Is it really better to transfer these tasks to a social planner instead of to the same decision maker? I understand that a priori our proposal might sound difficult to achieve, but what we are proposing is indeed what clinical psychologists and psychotherapists do every day. A therapeutic process begins by recognizing that there is a problem (point a), continues with trying to identify true preferences (point b) and finishes by finding a solution to it (point c) (see Hawton et al., 1989). Thus, if we want to understand the normative implications of Behavioural Economics, we should become familiar with the literature of Psychotherapy and clinical practice that studies the way in which people improve their decisions and life. This thesis provides a first step towards that aim.

As the reader may have noticed, the endogenous preference parameter is not restricted to be a frame, label or default option as in the above example. It can also be any psychological state like motivation, willpower or any endogenous situation or environment. Here is another example. Imagine yourself deciding whether to go to the gym or to stay at home watching TV. You are at home and your motivation to exercise is actually low. If you reflect enough to the extent that you are able to imagine yourself in the gym, you will be able to anticipate that once you are in the gym your motivation to exercise is high. A perfect reflective decision maker will characterize the object of choice as (gym, motivated to exercise) and (home, not motivated to exercise) and choose what his preferences over these two alternatives dictate. In contrast, a decision maker with imperfect reflection will not anticipate that his motivation change with him going to the gym, and therefore will evaluate the options of staying at home or going to the gym, from his perspective of a person with "low motivation to exercise." Reflection ensures a full characterization of the object of choice (gym and motivation), internalizing all its consequences.

In standard Economics, decision problems are presented as either assuming that preference parameters are purely exogenous or, otherwise, the assumption of perfect reflection ensures that the decision maker will know how to deal with the endogeneity. If any of these two assumptions held, we would be in the standard Economics world in which most of what I will say in this thesis would not make sense. However neither of these two assumptions is supported by empirical evidence.

The evidence in favour of the endogeneity of preference parameters is enormous. Most if not all psychological therapies are about identifying appropriate coping strategies to modify thoughts, beliefs, behaviour and environments to improve people's psychological states (see Hawton et al., 1989). People cope with stress, substance abuse, anger or anxiety, by changing their response to a situation (emotion-focused problem) or by changing the environment (problem-focused coping) (Lazarus and Folkman, 1984). Baron (2006, pg. 68) shows that emotions are (partly) under people's control and argues that individuals can "induce or suppress emotions in themselves almost on cue." Some people may even reshape their character, so that their emotional responses change. There is also extensive work in Social Psychology that views human functioning as the product of a dynamic interplay of personal, behavioural, and environmental influence. Albert Bandura, one of the main advocates of this view, points out that the way in which people interpret the results of their own behaviour informs and alters their environments and personal factors which, in turn, inform and alter subsequent behaviour through an "environmental feedback effect." He labelled this view "reciprocal determinism" (Bandura, 1986, 1997, 2001). Finally, the evidence of endogenous preferences parameters is not restrictive to Psychology. There is also evidence from Economics showing that dynamically inconsistent but sophisticated people actively manage their self-control problems by choosing activities that reduce the likelihood of encountering cues that trigger binges. For instance, people purchase costly pre-commitment devices like small packages of junk food even when the unit price is lower for larger packages (Wertenbroch, 1998) or self-impose costly deadlines to overcome procrastination (Ariely and Wertenbroch, 2002).

The evidence against the assumption of perfect reflection is more indirect than the evidence against exogeneity of preference parameters. The main point of our framework is that some people, in some situations, may fail to recognize that their preference parameters are endogenous and that they are eventually under their own control. We argue that this happens because people do not reach perfect reflection. Evidence of the existence of some systematic cognitive biases is enough as proof against perfect reflection. It has been shown that people misjudge self capacities to carry out certain actions (Baumeister and Scher, 1988), mispredict affective reactions overestimating regret and rejoicing (Sevdalis and Harvey, 2007) and fail to bear in mind that they will adapt (Fagerlin et al., 2005). Overall, the literature on affective forecasting shows that people make systematic inaccurate predictions about how they will feel in situations, even if the situation people experience objectively matches the situation they imagined. This proves reflection certainly imperfect.

Based on the above empirical grounds, this thesis develops a theoretical framework with endogenous preferences in which the assumption of perfect reflection is relaxed. We analyse the implications for both descriptive and Normative Economics and for public policy and introduce novel applications.

Chapter 2 introduces the main formal model of the thesis and studies its implications for Positive Economics. Our approach relies on two key premises: (a) preferences parameters are endogenously determined by choices and (b) the decision maker, due to imperfect reflection, may not fully internalize that effect. The degree of reflection of the decision maker will determine two types of decision problems. A "Standard Decision Problem" (SDP) which is one in which the decision maker is reflective enough to internalize the effect of his actions on his preference parameters and a "Behavioural Decision Problem" (BDP) when the decision maker reflects imperfectly and she does not internalize such effect. Importantly, a BDP assumes individual rationality in the sense that actions are required to be consistent with preferences parameters, and for fixed preference parameters the decision-maker chooses an optimal action. It is argued that the outcomes of a BDP correspond to Nash equilibria, and the outcomes of a SDP to the Stackelberg equilibria, of an appropriately constructed intra-self game.

The chapter proposes a new equilibrium existence result in pure actions without complete and transitive preferences. A result like that is important on its own, since incomplete and non-transitive preferences are a common token in Behavioural Economics models. Therefore, our framework not only allows the person to have imperfect reflection, but also to have incomplete knowledge of what she prefers.

It is shown that behavioural decisions are falsifiable, which respond to some concern in the literature that general behavioural models may not be falsifiable (see Green and Hojman, 2007). In addition, we show that standard and behavioural decision problems have different testable implications. This is an important step towards the answer of a more ambitious question on whether we could test if a person is solving a decision problem reflecting fully or not.

Next, it is shown that a BDP unifies seemingly disconnected models in the literature, from situations where the preference parameter corresponds to the decision maker's current state (Tversky and Kahneman, 1991; Shalev, 2000), beliefs (Geanakoplos et al., 1989; Akerlof and Dickens, 1982), emotions (Bracha and Brown, 2008), expected consumption (Koszegi, 2005; Koszegi and Rabin, 2006) or aspirations (Heifetz and Minelli, 2006), to models of adaptive preferences like Weizsäcker von (1971); Hammond (1976); Pollak (1978). All these papers can be classified as models with endogenous preferences parameters and some degree of imperfect internalization of the feedback effect[1].

Finally, the chapter shows that, for almost all classes of preferences, SDP and BDP are (outcome-wise) distinguishable from each other. The decision problems are indistinguishable when (a) the feedback map is a constant or (b) the utility function is additively separable in actions and preference parameters. In these two cases, relaxing the assumption of perfect reflection in a model with endogenous preferences is innocuous, i.e. it does not have any descriptive or normative implications. However, these are only isolated cases. We then explore some peculiar outcomes of distinguishable decision problems.

Next, in Chapter 3, we move on to study the normative implications of the framework proposed here. One of the dilemmas raised by Behavioural Economics is whether preferences inferred (or revealed) from choices can still be used as a valid welfare metric. Bernheim and Rangel (2009, 2008) and Rubinstein and Salant (2008) propose a normative framework that (a) relies only on observed choices and (b) encompasses situations in which choices are affected by frames or ancillary conditions. According to their welfare relation, every action chosen within some "welfare relevant choice domain" is a weak welfare optimum. The main drawback of their framework is that the set of weak welfare optima can be very thick and therefore it brings little information for public policies evaluation.

This thesis takes a different approach. As in standard Economics, we argue that the person has "true or normative" preferences which she satisfies under perfect reflection. In Friedman's (2003, p.5) words "self-reflection is the process in which, roughly, a whole self takes a stance toward particular wants and values she finds herself to have." With perfect reflection the person fully characterizes the objects of choice, ranks them using her "true preferences" and chooses the object that maximizes this ranking. With perfect reflection there is no wedge between "normative" and "revealed" preferences. However,

[1] Another related paper is Caplin and Leahy (2001), that introduces anxiety into an economic model.

if we allowed for imperfect reflection as we do here, a wedge may certainly emerge and challenges standard public Economics. We show that in general, "revealed preferences" from choices do not represent "normative preferences." Nevertheless, we offer the necessary and sufficient conditions under which, regardless the degree of reflection, "revealed preferences" over actions are valid to construct welfare rankings. These conditions are too strong though.

As we suggested in the beginning of this introduction, a perfect reflective person must (a) acknowledge the existence of a feedback map (i.e. that is acknowledging that his or her preference parameter is not exogenous and she can do something to affect it), (b) identify his or her coping strategies (the set of actions that can be taken to affect the preferences parameters), (c) identify how such strategies affect his or her preference parameters (fully identification of the feedback map) and (d) implement her strategies efficiently. In that sense, perfect reflection is equivalent to the concept of "autonomy" studied in the literature of Philosophy and Psychology. Deeper reflection implies deeper autonomy. As argued by philosopher Friedman (2003, pg. 4) "autonomous behaviour is based on the deeper wants and commitments of the person, is partly caused by her reflections on and reaffirmations of them." To realize autonomy a person must first somehow reflects on her wants and takes up an evaluative stance with respect to them. Friedman (2003, pg. 4) makes very clear the endogeneity of preference parameters when she argues that "for choices and actions to be autonomous, the choosing and acting self as the particular self she is must play a role in determining them [...] When wants and desires lead to choice or action without having been self-reflectively endorsed by the person, the resulting choices and actions are not autonomous." By the same token, Elster (1983, pg. 22) defines autonomous preferences (or desires) as those "that have been deliberately chosen, acquired or modified – either by an act of will or by a process of character planning".

In the intra-self game-theoretic interpretation of our framework, being autonomous can be understood as playing a la "Stackelberg" and being non-autonomous as playing a la "Cournot." So, what matters for welfare analysis in our framework is how the person plays her intra-self game. If the person plays her game a la Stackelberg, then the revealed preferences from choices will be equivalent to her normative preferences, and therefore, can guide normative judgments.

Overall, this thesis provides a formal framework in which the degree of personal "autonomy" emerges as a natural normative criterion.

We assess the autonomy criterion in light of the other existing non-choice base normative criteria such as opportunity (Sugden, 2004) or individual happiness (Kahneman et al., 1997). We argue that the autonomy criterion encompasses the above mentioned criteria in the following sense: if the person is autonomous (a) her choices will reveal her true preferences, (b) more opportunities will never be bad and (c) her subjective well-being may be a valid measure of her welfare. If the person is not autonomous, then any of these criteria may not be appropriate as a normative criterion.

Then, we undertake a careful comparison of our framework with the choice theoretic models of Bernheim and Rangel (2009, 2008) and Rubinstein and Salant (2008) in which observed choices are determined by frames or ancillary conditions. The main concern of these models is whether such frames or ancillary conditions matter from a welfare viewpoint. There are at least two diametrically opposite views: (a) the choices of an autonomous decision maker will not depend on frames or ancillary conditions and therefore when frames and ancillary conditions do affect choices they could, in principle, matter for welfare, and (b) frames or ancillary conditions are exogenous to choices and are welfare irrelevant from a social planner's viewpoint.

If the view is (a), we argue that the choice mappings studied in BR and RS correspond to a decision problem where each frame or ancillary condition is consistent with any action chosen by the decision-maker. In a BDP, the decision-maker treats the frame or ancillary condition as given while in a SDP frames or ancillary conditions are considered to be characteristics of actions and are fully taken into account. We show that the set of outcomes of a SDP is a subset of the set of outcomes of a BDP and the set of welfare optima derived solely from observed choice contains all the actions corresponding to the solution of a SDP. For a general feedback map from actions to preference parameters, we show that the weak welfare optima derived from observed choice may have an empty intersection with, or exclude elements of, the set of actions that correspond to the outcomes of a SDP.

If, on the other hand, both frames and ancillary conditions are truly exogenous to choices and welfare irrelevant for a social planner (case (b)), we argue that the choice mappings studied in BR and RS are generated in a family of distinct decision problems each indexed by a different preference parameter. In this case, the feedback from actions to preference parameters is a constant, the outcomes of BDP and SDP coincide and the set of welfare optima derived solely from observed choice is exactly the actions corresponding to the solution of a SDP.

At the end of chapter 3, we analyze some policy implications of our argument. Clearly, when the outcomes of both BDP and SDP coincide, there is no case for any sort of intervention by a social planner. In contrast, in scenarios where the set of outcomes of a SDP is contained in the set of outcomes of a BDP, we argue that the "libertarian paternalism" approach to policy interventions, advocated by Sunstein and Thaler (2008), that only seeks to alter the frames or ancillary conditions of individuals might work. However, we finish the chapter of normative implications arguing that policy interventions should, in general, aim to ensure that decision-makers internalize the feedback from actions to preference parameters.

So far, we have argued that the assumption of perfect reflection in an individual decision-making setting is strong. This assumption becomes even stronger in a N-person strategic interaction setting. With this motivation in mind, chapter 4 extends the individual decision framework to a game. A normal-form game with endogenous preference parameters and players who do not fully internalize the effect of their actions is labelled a "psycho-social game."

Most of the interesting theoretical aspects of "psycho-social games" have been explored in the one-person setting introduced in chapter 2. Therefore, the focus of this chapter is mostly on applications. We link our framework to the existent literature on social preferences and develop two applications of psycho-social games introducing the concept of individual commitment and empowerment into game theory.

On the theoretical front, we first prove existence under weak assumptions, extending the existence proof of Chapter 2. We associate psycho-social games to standard normal-form games in two different ways and show that, typically, the set of Nash equilibria and the set of psycho-social equilibria of an associated psycho-social game are distinct from each other. We also show that a strict Nash equilibrium is robust to any kind of imperfect reflection. This result is consistent with Aumann's (1997) claim that "no rationality at all is required to arrive at a Nash equilibrium" (pg. 4). This is true when the normal-form game has a unique strict pure strategy Nash equilibrium. If the game has weak or mixed strategy equilibria, then we show that some degree of "rationality" in Elster's (1983) broad sense is needed to sustain a weak Nash equilibrium.

Chapter 5 applies our framework to study the interaction between people's aspirations and their initial extrinsic disadvantages. Debraj Ray once wrote "poverty stifles dreams" (2003, pg. 1). This painful but, in my view, real statement was highly inspirational. As an economist, I thought: if poverty affects dreams or aspirations and we model them as preference parameters, then we would not be able to make sense of Ray's statement in a classic microeconomics model in which endowments and preferences are orthogonal to each other. Moreover, suppose that we had a model in which endowments affect preferences. Would this effect be permanent or temporary? In other words, are dreams stifled forever? or dreams would flourish as soon as an external policy (e.g. relaxing credit constraints) suddenly placed the poor "above the poverty line". My first impression when I started this Ph.D. was that poverty is a much more complex concept than just being below or above an income threshold. This impression was confirmed when I became familiar with the multidisciplinary literature covering the topic. Poverty is a relative condition that is "fundamentally self-perpetuating [...] and goes hand in hand with lack of hope" Mookherjee and Ray (2003, pg. 5). So the effect of initial disadvantage on psychological states like self-confidence, aspirations and hope may possibly not be changed only by altering the external conditions that brought the individual to that self-fulfilling equilibrium, but mainly by relaxing those psychological states that are becoming an internal constraint to the individual. Empowerment policies have this aim.

This phenomenon is studied in the literature of Psychology with the name "learned helplessness". The principle behind learned helplessness is that the individual infers on the basis of stress or unhappy experiences of her or relevant others that she is incapable of exerting control of his life and achieving desirable outcomes (Baumeister and Scher, 1988). The important thing is that such expectations or attributions about her capabilities turn out to be maladaptive and become part of the person's identity. Due to this effect, initial disadvantages may affect the autonomy of the individuals.

Chapter 6 provides a modest first step to model this complex idea. We consider a decision-maker who is endowed with an initial status and who has to choose between two actions: one that perpetuates her initial status and one that has the potential of improving her initial status. Otherwise, the decision-maker chooses an action that has the potential of improving her status, she will reach her target with probability p and stays with her initial status with probability 1-p where p is a combination of external signals (e.g. a role model) and her subjective beliefs (e.g. self-confidence). The decision maker gets a benefit from the final status achieved and there is a cost of action which depends on her initial status. Choosing the action that perpetuates the status-quo yields no benefit and is costless. Any attempt to change the status-quo is costly but (a) the cost of doing so is lower the higher is the initial status (so that the initial status is a measure of the level of disadvantage of the decision-maker) and (b) the target status is achieved with probability p and the benefit function is increasing in the target status (and hence, the aspiration level).

We first show that initial advantage and self-confidence are substitutes for aspirations. That is, the more disadvantaged the decision-maker is (respectively the lower self-confidence), the higher aspirations she needs in order to find it convenient to undertake an action that has the potential to change her status-quo. The main implication of this result is that initial disadvantage is associated with low aspirations and low achievement.

Next, we allow for endogenous aspirations. We study the impact of a role model as a way out of a deprivation trap. A role model provides direct information about p through a successful experience of a "similar" other. The more similar the role model is to the decision-maker the more likely will it be to break an aspiration failure. People choose the sample from which to extract a role model, a so called “cognitive neighbourhood”. It is shown that an initially disadvantaged decision-maker will find it optimal to restrict the size of her "cognitive neighbourhood" to those who are similar to her. We use this result to construct an example to show that low aspirations and low achievement arise in a polarized society. Finally, we allow for aspirations to be adaptively determined by choices, applying the model introduced in Chaper 2. We show that for some medium level of initial circumstances, a policy that increase the autonomy of the individual creates a big positive change in his condition, without having to change her external condition exogenously.

In short, Chapter 5 studies how initial circumstances and behaviour affect aspirations. The last chapter, instead, takes initial circumstances and aspirations as given and explores a novel channel through which aspirations affect behaviour. Precisely, Chapter 6 enquires into what extent people’s real life income aspirations affect their level of cooperation in general and, in particular, with whom they cooperate.

The intuition behind the first part of the enquiry, that is, income aspirations affecting people's cooperation is as follows. We conjecture that the marginal rate of substitution between individual and collective welfare is decreasing on the extent to which individual's income aspiration is satisfied. Importantly, note that this statement is different to saying that individual and collective welfare are somehow complementary. Our statement implies a one-way causality consistent with some implicit quasi-lexicographic order between the two dimensions (individual welfare and others' welfare). We want to investigate whether people care first about satisfying some material needs or aspirations, and as long as those needs are relatively satisfied, they start caring about the welfare or needs of the others. Implicit in this statement, it is the view that people are not intrinsically either selfish or altruist as the existing literature seems to suggest, but their level of cooperation depends on how satisfied their material aspirations are. This hypothesis is grounded on two orthogonal and well-known theories. One is Simon's (1955) idea that people look for achieving satisfactory targets instead of maximizing utility. The other is Maslow's (1943) theory of hierarchy of needs. In this sense, individual material aspirations would correspond to the lower steps of Maslow's pyramid and social welfare can be understood as the upper levels of the pyramid. The two theories support the hypothesis that the marginal rate of substitution between both dimensions depends on the extent to which the first dimension is satisfied.

In addition to this hypothesis, we investigate whether geographical identity of the recipients of a public goods (neighbours, compatriots or foreigners) affect the propensity to cooperate.

By combining experimental data with a post-experimental survey we find evidence that the more frustrated people are with their income, the lower is their propensity to cooperate to provide global (e.g. environmental) and national public goods. We show that it is the gap between aspirations and actual income that matters, and not the absolute value of income or aspirations. The quantitative effect is considerably high. Participants who are most frustrated are 46 percent more likely to free-ride on foreigners than those who are satisfied with their income. As a parallel result, we found that the effect of group identity on cooperation depends on whether the public good is local, national or global. Contrary to what is assumed in the literature on social identity and cooperation, in-group identification with compatriots does not affect participants' contributions to a national public good, although it does affect contributions to local and global public goods.

My hope is that, if I managed to convince you that it is worth having a look at this thesis, you will enjoy it. This piece of work is just a snap-shot of what I have investigated during this Ph.D. journey and by no means have all my questions been answered. Many of them are still pending and others have now been opened as a result of this thesis. I leave chapter 7 to highlight some of the most challenging ones.

# Chapter 2

# Behavioural Decisions: Positive Analysis

## 2.1 Introduction

This chapter develops the main theoretical framework of the thesis and studies its implications for descriptive Economics. Our approach relies on two key premises: (a) preferences parameters[1] are endogenously determined by choices and (b) the decision maker, due to imperfect reflection, may not fully internalize that effect. The degree of reflection of the decision maker defines the procedure of choice. We consider two rather extreme types of decision problems. A "Standard Decision Problem" (SDP) which is one in which the decision maker is reflective enough to internalize the effect of his actions on his preference parameters and a "Behavioural Decision Problem" (BDP) when the decision maker reflects imperfectly and she does not internalize such effect[2].

Our model can be considered as a reduced-form representation of an intra-personal

[1]Throughout the thesis we use "psychological states" or "preference parameters" interchangeably. They should be interpreted as any pay-off relevant parameter that can eventually be affected by the choice of the individual. Mood, beliefs, aspirations, attitudes, emotions or values are some examples.

[2]We study two extreme cases to gain theoretical understanding of the entire framework. We acknowledge, though, that many "realistic" and maybe interesting situations are somewhere between both decision problems. We consider this case in the public policy section in Chapter 3.

game between two selves. A "conscious" self-1, who chooses actions to maximize his utility, and an "unconscious" self-2, who chooses psychological states to minimize some deeper psychological needs like anxiety, frustration or fear. The Nash equilibria of this intra-personal game are equivalent to the outcomes of a BDP, while the Stackelberg equilibrium is equivalent to the outcome of a SDP. A person who solves her intra-personal game *a la* Stackelberg is autonomous[3], that is, she is in control of her own psychological states. A non-autonomous person, in contrast, is the person who solves her own intra-personal game *a la* Nash, choosing an action taking the psychological states of self-2 as given[4].

The chapter begins by introducing a simple example of reference-dependent preferences (Tversky and Kahneman, 1991) which highlights, in a simple and intuitive way, the key aspects of our general framework in a simple and intuitive way and it prefigures what we do in the rest of the chapter.

Section 2.3 introduces the general formal framework. Theorem 1 (sub-section 2.3.1) proposes a new equilibrium existence result in pure actions without complete and transitive preferences. A result like that is important on its own, since incomplete and non-transitive preferences are a common token in Behavioural Economics models. Therefore, our framework not only allows the person to have imperfect reflection, but also to have incomplete knowledge of what she prefers. In addition, sub-section 2.3.2 studies the testable restrictions of our theory. Proposition 1 shows that the behavioural decisions are falsifiable, which respond to some concern in the literature that general behavioural models may not be falsifiable (see Green and Hojman, 2007). It is shown that standard and behavioural decision problems have different testable implications.

Section 2.4 shows that BDPs unify seemingly disconnected positive Behavioural Eco-

[3]Autonomy is the regulation by the self (Ryan and Deci, 2006). In Greek: "Auto-Nomos" - nomos meaning "law": one who gives oneself his/her own law. We review this concept more in detail in the last section of this chapter and notably in chapter 3.

[4]For the rest of the paper, we refer to the person not having control of her psychological states as equivalent to the person taking her preference paramenters or psychological states as given.

nomics models. This make BDPs a natural framework to study general properties of Behavioural Economics models.

Next, we devote section 2.5 to the analysis of indistinguishability. Proposition 2 introduces the necessary and sufficient conditions over preferences for SDP and BDP to be (outcome-wise) indistinguishable from each other. Theorem 2 shows that for almost all classes of preferences, SDP and BDP are (outcome-wise) distinguishable from each other. In particular, decision problems are indistinguishable from each other when (a) the feedback map is a constant or (b) the utility function is additively separable in actions and preference parameters. In these two cases, relaxing the assumption of perfect reflection in a model with endogenous preferences is innocuous, i.e. it does not have descriptive or normative implications. Section 2.5.3 explores some peculiar outcomes of distinguishable decision problems.

In section 2.6 we show an example that illustrates how to introduce the notion of "coping strategies" into a general individual decision-making framework. Our account includes models of dynamic inconsistent preferences and self-control problems (Bernheim and Rangel, 2004, 2005) but by all means it is not restricted to them.

The chapter ends with a short review of relevant literature in Philosophy and Psychology that supports our theoretical framework.

## 2.2 Example: Reference-Dependent Preferences

This example has two aims. It highlights the key aspects of our general framework in a simple and intuitive way and it prefigures what we do in the rest of the chapter.

Consider a person who has to decide whether to stay in the status-quo, $q$, or to change, $a$. The cost of changing is $c$. The person derives utility from a $K$-dimensional bundle of outcomes, $x$, $u(x) = \sum_k u_k(x_k)$.The utility that the person derives from the decision to change, $u(a)$, depends on the state of the world, $s = 1, ..., S$ with a probability

$\mathbf{p}_s$ associated to each state of the world. The expected utility is denoted by $U(x) = \sum_s \mathbf{p}_s u(x)$. A person will change her status-quo if and only if the expected pay-off of changing is higher than the pay-off from remaining in the status-quo, $U(a) > U(q)$:

$$\sum_s \mathbf{p}_s u(a) - c > u(q) \tag{2.1}$$

The policy implications of this problem are well-known in standard Economics. A change in prices, information, utility parameters (e.g. risk aversion) or the cost of changing will affect individual behaviour.

Now consider the same problem in a context of a behavioural model with reference-dependent preferences (Tversky and Kahneman, 1991). The pay-off of the person is given by

$$u(x,r) = \sum_k \left[u_k(x_k) + v_k(x_k - r_k)\right], \tag{2.2}$$

where $v_k(x_k - r_k)$ is a value function defined over gains and losses relative to a reference point $r_k$ as introduced by Kahneman and Tversky (1979). Think of the first part of the utility function $\sum_k u_k(x_k)$ as the "outcome-based utility" usually studied in Economics and the second part $\sum_k v_k(x_k - r_k)$ as the reference-dependent "gain-loss utility." Suppose $v(x) = x$ if $x \geq 0$, $v(x) = \alpha x$, $\alpha > 1$ if $x < 0$ and $v(0) = 0$. Since $\alpha > 1$, $v(x)$ is steeper in the loss side, so this functional form captures the idea of loss-aversion: losses (or outcomes below the reference point) loom larger than corresponding gains (or outcomes above the reference point).

For expository purposes, suppose there are only two dimensions of choice, $K = 2$[5]. The status-quo option is defined by $q = (q_1, q_2)$ and the alternative option is $a = (a_1, a_2)$,

[5] A tipical example of two possible dimensions are mugs and money.

$q_1 > a_1$ and $q_2 < a_2$. As it happens with any reference-dependent theory, its predictions depend crucially on what determines the reference point. One of the possible avenues is to equate the reference point with expectations (Koszegi and Rabin, 2006). If the person expects to maintain the status-quo (respectively to change), her reference point will be equated to the status-quo (respectively to the alternative). In this way, the reference point or expectations is endogenously determined by the identity map $\pi : \{q, a\} \rightarrow \{q, a\}$. To complete this example, suppose the person has rational expectations so that she is able to predict her own behavior[6]. Thus, the only two possible outcomes (in pure actions) of this decision problem are $(a, r = a)$ and $(q, r = q)$.

If the person has expectations to remain in the status-quo, her reference point will be $r = q$ and given that, she will chose to remain in the status-quo if and only $U(q, r = q) > U(a, r = q)$:

$$\sum_k u_k(q_k) > \sum_s p_s \left[ \sum_k u_k(a_k) + v_k(a_k - q_k) \right] - c \tag{2.3}$$

In which case $(q, r = q)$ is a behavioural outcome.

Alternatively, if the person expects to change, she will use $r = a$ as a reference point, and given that, she will chose to change if and only if $U(a, r = a) > U(q, r = a)$:

$$\sum_k [u_k(q_k) + v_k(q_k - a_k)] < \sum_s p_s \left[ \sum_k u_k(a_k) \right] - c \tag{2.4}$$

In which case $(a, r = a)$ is a behavioural outcome.

In contrast to the unique outcome of the problem of standard Economics with which we started the example, this problem with reference-dependent preferences has two behavioural outcomes: remaining in the status-quo expecting to do so $(q, r = q)$ and changing the status-quo expecting to do so $(a, r = a)$. More importantly, note that the

[6] As Koszegi and Rabin (2006) we rule out situations where the action implied by optimal behaviour conditional on expectations does not coincides with expectactions.

outcome $(a, r = a)$ welfare-dominates $(q, r = q)$ if and only if $U(a, r = a) > U(q, r = q)$:

$$\sum_s p_s \left[ \sum_k u_k(a_k) \right] - c > \sum_k u_k(q_k) \tag{2.5}$$

Note that condition 2.5 is equivalent to condition 2.1: $U(a, r = a) > U(q, r = q) \Leftrightarrow U(a) > U(q)$. Therefore, if the person could eventually choose her reference point, she would choose $r = a$ and the *positive* and *normative* implications of the behavioural decision problem would be equivalent to those of the standard decision problem.

An important message from this simple example is that this person would be (at least weakly) better-off if she internalized the feedback effect from her actions to the reference point. However, we keep the welfare analysis for next chapter, and we devote this chapter to study the descriptive implications of the general framework.

### 2.2.1 Individual Decision Problem as an Intra-personal Game

Behind the model of endogenous reference-dependent preferences presented above, there is an intra-personal game between two different selves, self-1 and self-2. Self-1 chooses $x \in X$ and self-2 chooses $r \in R$. Self-1, can be thought to be the "conscious" part of our selves, the one that a third person observes making choices. Self-2, can be interpreted as the "unconscious" part of the person, the one that aims to satisfy deeper inner psychological needs such as anxiety, fear or peace of mind. For instance, self-2 may prefer to expect to remain in the status-quo in order to minimize frustration, to believe that her job is safe to reduce fear, or to be an optimist to minimize anxiety. Each self has preferences over actions and reference-points, which may or may not conflict with each other. In the context of our example of above, preferences of self-1 are represented by the reference-dependent expected utility:

$$U(x, r) = \sum_s \mathbf{p}_s \sum_k [u_k(x_k) + v_k(x_k - r_k)] - c \tag{2.6}$$

and preferences of self-2 can be thought of being a cost function that self-2 aims to minimize choosing a reference-point $r \in R$. For example:

$$c_2(x, r) = \frac{1}{2}r^2 - r.x \tag{2.7}$$

One can interpret self-2's "preferences" as if she looks for a balance between expectations and actions. We can think that self-2 does not like feeling frustrated if she mispredicts her actions[7]. Formally, self-2's best reply is $\pi(x) = \arg\min_{r' \in R} c_2(x, r')$. Consistent with the assumption of rational expectations imposed in the example, self-2 best reply is the identity map $\pi(x) = x$.

Therefore, our example can be modeled as an intra-self game in which the best reply function of self-1 is

$$\alpha(r) = \begin{cases} q \text{ whenever } r = q, \\ a \text{ whenever } r = a, \end{cases} \tag{2.8}$$

and the best reply function of self-2 is $\pi(x) = x$ for each $x \in X$.

If the decision maker plays this intra-self game as a simultaneous move game, the Nash equilibrium is a decision state $(x^*, r^*)$ for which $x^* \in \alpha(r)$ and $r^* = \pi(x^*)$. This is exactly the definition of a behavioural equilibrium which will be formally introduced in next section. However, if one of the selves is allowed to control the other self's decisions, then the appropriate solution concept is Subgame Perfect Equilibrium. Suppose that the decision maker solves her intra-self game as a "Stackelberg leader." In our example, she will anticipate that if she plays $a$ her self-2 would play $r = a$. Since $(a, r = a) = \text{argmax}_{x \in X} u(x, \pi(x))$, $(a, r = a)$ is a sub-game perfect equilibrium of the intra-personal game or the "Stackelberg equilibrium." Note that $(a, r = a)$ is also the unique outcome of the Standard Decision Problem in our example.

[7] This could be interpreted as a psychological foundation for rational expectations: self-2's best response motivated by psychological needs makes the person chooses what she expects to have.

### 2.2.2 Remarks from the Example

The above simple example of reference-dependent preferences highlights several important features that motivate the rest of this chapter.

**(a) Theoretical implications**

**(a.1) Synthesis:** In the example above, the reference points are people's expectations. More generally, $r$ can be any pay-off relevant preference parameter that can eventually be affected by people's will. This include mood, points of view, beliefs, aspirations, attitudes, emotions or values. In this sense, as we show in Sub-section 2.4, this model becomes a reduced-form representation of seemingly disconnected models in the literature, each model assuming a different $X, R$, $u$ and some implicit feedback $\pi$.

**(a.2.) Intra-personal normal-form game:** In the same way we showed in our example, the vast range of models that are reduced to our framework can also be modeled as an appropriate intra-self normal-form game between two selves, one choosing actions and the other choosing preference parameters. Whether the decision maker solves her intra-self game *a la* Nash or *a la* Stackelberg has important positive and normative implications.

**(a.3.) Adaptive dynamics:** although it is not explicit in the example, it can be easily shown that a behavioural equilibrium or a Nash equilibrium of an intra-self game is the steady state of an adaptive dynamics over actions and psychological states. A behavioural equilibrium is reached after people adapt their reference-point to their actions[8].

**(c) Positive implications:** the way in which the person solves her intra-self game may generate different economic predictions. We have learned that the outcome

[8]This may not happen in pure actions though. We allow also for counteradaptive preferences which is the opposite phenomenon of adaptive preferences (Elster, 1983, pg. 111). Example 5 illustrates this point.

$(q, q)$,given the preferences assumed in the example, would never be an outcome of a SDP. In such case, the two decision problems are "theoretically" distinguishable from each other. However, under some conditions over preferences, it may not matter the degree of reflection of the decision-maker. In Section 2.5, we offer the general conditions on preferences under which any two decision problems with the same $X, R, u$ and $\pi$ are distinguishable from each other.

**(d) Testable implications:** our framework opens new questions regarding identification of the decision process. Suppose we only observe actions chosen for any possible sub-set of actions. Can we identify whether the person is being perfectly reflective or not? In Section 2.3, we show that both decision problems have different testable implications. That is, there are data that are behavioural-rationalizable and not standard-rationalizable, every outcome that is standard-rationalizable is also behavioral-rationalizable, and there are outcomes that are not behavioural-rationalizable.

**(d) Normative implications:** A status-quo bias $(q, q)$ would not be a possible outcome of a model in which preference parameters were endogenous and decision-makers were assumed to be perfectly reflective. Therefore, the assumption on how the decision-maker plays her intra-personal game is crucial for normative analysis. More importantly, the revealed preference for choosing the status-quo over changing it when the person is trapped in a status-quo bias is not an appropriate indicator for welfare. In Chapter 3 we address this issue and characterize the general conditions on primitive preferences under which revealed preferences can be recovered as an appropriate welfare indicator.

**(e) Policy implications:** this example already pins down the fact that the existing public policies of standard Economics may not be sufficient. In the case where there is a status-quo bias, one could consider three distinctive policies: (a) to affect $r$ exogenously, (b) to affect the expectations formation itself or (c) to affect the decision process enhancing the abilities of the individual to choose the optimal reference point. In economics terms, this latter policy can be interpreted as learning how to play the intra-personal game as a Stackelberg leader. In philosophical and psychological terms, it can be interpreted as gaining individual autonomy. We discuss these concerns in Chapter 3.

The rest of the chapter introduces the general model and studies in a general way the insights highlighted by this example.

## 2.3 The General Framework

A (complex) decision scenario $D = (A, P, \pi, \succeq)$ consists of a set $A \subset \Re^k$ of actions, a set $P \subset \Re^n$ of psychological states[9], a mapping $\pi : A \to P$ modelling the feedback effect from actions to psychological states and a binary relation $\succeq$ ranking pairs of $(a, p)$ in $(A \times P) \times (A \times P)$. It is assumed that $\pi(a)$ is single-valued and non-empty for each $a \in A$, and $\Re^k$ and $\Re^n$ are finite dimensional Euclidian spaces. The binary relation $\succeq$ assumes that the individuals are not only able to rank actions given a psychological state (i.e. $(a, p) \succeq (a', p)$), but also that they are able to rank psychological states for a given action (i.e. $(a, p) \succeq (a, p')$). This assumption is analogous to the one made by Harsanyi (1953).

A decision state is a pair of action and psychological state $(a, p)$ where $a \in A$ and $p \in P$.

The preferences of the decision-maker are denoted by $\succeq$. The expression $\{(a, p), (a', p')\} \in \succeq$ is written as $(a, p) \succeq (a', p')$ and is to be read as "$(a, p)$ is weakly preferred to (equivalently, weakly welfare dominates) $(a', p')$ by the decision-maker".

A *consistent state* is a decision state $(a, p)$ such that $p = \pi(a)$. Let

$$\pi(A) = \{p \in P : \exists\, a \in A \text{ s.t. } p = \pi(a)\}$$

be the set of consistent psychological states, and

[9] I decided to call $p$ psychological state, but it can be interpreted more broadly as any preference parameter that is potentially affected by individual behaviour.

$$\Omega = \{(a,p) \in (A \times P) : p = \pi(a) \text{ for all } a \in A\}$$

be the set of consistent decision states.

There are two types of decision problems studied here:

1. A *standard decision problem* ($SDP$) is one where the decision-maker chooses a pair $(a,p)$ within the set of consistent decision states. The outcomes of a $SDP$ are denoted by $M$ where

$$M = \{(a,p) \in \Omega : (a,p) \succeq (a',p') \text{ for all } (a',p') \in \Omega\}.$$

2. A *behavioral decision problem* ($BDP$) is one where the decision maker takes as given the psychological state $p$ when choosing $a$. Define a preference relation $\succeq_p$ over $A$ as follows:

$$a \succeq_p a' \Leftrightarrow (a,p) \succeq (a',p) \text{ for } p \in P.$$

The outcomes of a $BDP$ are denoted by $E$ where

$$E = \{(a,p) \in \Omega : a \succeq_p a' \text{ for all } a' \in A\}.$$

Suppose $P = A$ and $a = \pi(a)$. In this case, the decision problems studied here offer a way of modelling situations where "the reference state usually corresponds to the decision maker's current state." (Tversky and Kahneman, 1991, pg. 1046). The following examples show that whether the decision-maker correctly anticipates the feedback effect from actions to the preference parameter or not, will have an impact on the decision outcomes.

**Example 1** *($M \subset E$)*

Consider a decision problem where $A = P = \{a_1, a_2\}$, $\pi(a_i) = \{a_i\}, i = 1, 2,$

and $(a_i, a_i) \succ (a_j, a_i)$, $j \neq i$ and $(a_1, a_1) \succ (a_2, a_2)$. Then, $M = \{(a_1, a_1)\}$ but $E = \{(a_1, a_1), (a_2, a_2)\}$.

**Example 2** *($M \neq \varnothing$, $E \neq \varnothing$, $M \cap E = \varnothing$)*

Consider a decision problem where $A = P = \{a_1, a_2\}$, $\pi(a_i) = \{a_i\}$, $i = 1, 2$, and $(a_2, a_j) \succ (a_1, a_j)$, $j = 1, 2$, and $(a_1, a_1) \succ (a_2, a_2)$. In words, $a_2$ is preferred in all states and it is better to be in state $a_1$.Then, $M = \{(a_1, a_1)\}$ but $E = \{(a_2, a_2)\}$.

**Example 3** *($M \neq \varnothing$, $E = \varnothing$)*

Consider a decision problem where $A = P = \{a_1, a_2\}$, $\pi(a_i) = \{a_i\}$, $i = 1, 2$, and $(a_j, a_i) \succ (a_i, a_i)$, $i \neq j$, and $(a_1, a_1) \succ (a_2, a_2)$. Then, $M = \{(a_1, a_1)\}$ but $E$ is empty.

### 2.3.1 Existence

It is not hard to check that as long as both $A$ and $P$ are finite and $\pi(a)$ is single-valued for each $a \in A$, a random equilibrium exists, i.e. $E$ is not empty. Instead, this chapter studies existence in situations where the underlying preferences are not necessarily complete or transitive and underlying action sets are not necessarily convex. Mandler (2005) shows that incomplete preferences and intransitivity is required for "status quo maintenance" (encompassing endowment effects, loss aversion and willingness to pay-willingness to accept diversity) to be outcome rational. Tversky and Kahneman (1991) argue that reference-dependent preferences may not be convex. So we allow preferences to be incomplete, non-convex and acyclic (not necessarily transitive) and we show existence of a behavioural equilibrium in pure actions extending Ghosal's (2006) result for normal-form games to behavioural decision problems[10].

[10] The seminal proof for existence of equilibria with incomplete preferences in Shafer and Sonnenschein (1975) requires convexity both for showing the existence of an optimal choice and using Kakutani's fixpoint theorem.

**Theorem 1** ***(Existence)*** *Suppose $\pi : A \to P$ is increasing in $A$. Under assumptions of single-crossing, quasi-supermodularity and monotone closure, a pure action behavioral equilibrium exists.*

### 2.3.2 Testability

Our model is about two distinctive theories of individual behaviour: one characterized as a Standard Decision Problem ($SDP$) and the other as a Behavioral Decision Problem ($BDP$). Are these theories falsifiable? If so, are the testable implications of each theory different from each other? Below we show that the answer to these questions is Yes, they are falsifiable and have different testable implications.

A theory is falsifiable if there exists some outcome that cannot be rationalized as an equilibrium of that theory. For example, standard choice theory is falsifiable if Arrow's (1959) choice axiom holds[11]. Arrow's choice axiom states that when the set of feasible alternatives shrinks, the choice from the smaller set consists precisely of those alternatives that were selected from the larger set and remain feasible, if there are any. What can we say about our two theories?

Take $u : A \times P \to \Re$ as a numerical representation of $\succeq$, $\pi : A \to P$ and a family $\tilde{\mathbf{A}}$ of subsets of $A$. Define two correspondences, $\hat{c}$ and $\tilde{c}$, from $\tilde{\mathbf{A}}$ to $A$ as

$$\hat{c}(A') = \text{argmax}_{a \in A'} u(a, \pi(a))$$

and

$$\tilde{c}(A') = \{a^* : a^* \in \text{argmax}_{a \in A'} u(a, \pi(a^*)\},$$

so, the choices of the standard (autonomous) and behavioural (non-autonomous) person, respectively.

Assume that $A$, $P$ and all $A' \in \tilde{\mathbf{A}}$ are convex subsets in $\Re$. If we assume that

[11] For an example of a general behavioural model that is not falsibiable see Green and Hojman (2007).

$v(a) = u(a, \pi(a))$ is strictly concave, then $\hat{c}(A')$ is single-valued. On the other hand, if $A = P = [0, 1]$, $\pi(a) = a$ and

$$u(a, p) = -(a - p)^2 - (1/2 - p)^2, \tag{2.9}$$

then $v$ is strictly concave and $\hat{c}(A') = \{1/2\}$, but $\tilde{c}(A') = A$.

Then, the concern is that the explanatory power of $\tilde{c}$ is null. Suppose that we observe a correspondence $H$ from $\tilde{\mathbf{A}}$ to $A$ such that $H(A') \subseteq A'$. One may want to test two type of null hypothesis, one weak and the other strong. A weak null hypothesis for SDP would be that there exist $P$, $\pi$ and $u$ such that $v$ is strictly concave and $H(A') \subseteq \hat{c}(A')$. Under this null, it must be that $H$ is single-valued and satisfies the strong axiom of revealed preferences, i.e. SDPs have strong restrictions. However, if we want to test the weak null hypothesis for BDP that there exist $P$, $\pi$ and $u$ such that $v$ is strictly concave and $H(A') \subseteq \tilde{c}(A')$, then a BDP does not seem to be falsifiable.

This latter observation does not apply if we impose a stronger hypothesis and require that $H(A') = \tilde{c}(A')$[12]. So suppose that we want to test this strong hypothesis and suppose that $H$ captures all the choices of the person. We know that under the null that there exist $P$, $\pi$ and $u$ such that $H(A') = \hat{c}(A')$, we have that

**C.3:** If $A'' \subseteq A'$ and $a \in H(A') \cap A''$ then $a \in H(A'')$, and

**C.4:** If $\{a, a'\} \subseteq H(A'') \cap A'$ and $a \in H(A')$ then $a' \in H(A')$

Condition **C.3** is analogous to the Independence of Irrelevant Alternatives Axiom and Condition **C.4** to the Weak Congruence Axiom[13]. Under the null that there exist $P$, $\pi$ and $u$ such that $H(A') = \tilde{c}(A')$, we have that **C.3** holds but **C.4** doesn't hold. **C.3** holds

[12] The cost of testing this stronger null hypothesis is that we have to assume that we observe in our data all the points in $\tilde{c}(A')$ when we apply the test. Note that this assumption is not needed in the first case, since the null hypothesis guarantees that $\hat{c}(A')$ is single-valued, so if we observe one point of $\hat{c}(A')$, under the null, we are observing all the points.

[13] (see Sen, 1971)

because if $a \in \text{argmax}_{a' \in A'} u(a', \pi(a))$ and $a \in A'' \subseteq A'$, then $a \in \text{argmax}_{a' \in A''} u(a', \pi(a))$. To show that **C.4** doesn't hold if the null is strong, consider the following example: $A = \{a_1, a_2, a_3\}$, $P = \{p_1, p_2\}$, $\pi(a_1) = \pi(a_3) = p_1$, $\pi(a_2) = p_2$, and $u(a, p)$ is:

| $u(a,p)$ | $p_1$ | $p_2$ |
|---|---|---|
| $a_1$ | 3 | 1 |
| $a_2$ | 2 | 2 |
| $a_3$ | 1 | 3 |

(2.10)

In this case, $\tilde{c}(A) = \{a_1\}$ but $\tilde{c}(\{a_1, a_2\}) = \{a_1, a_2\}$.

We can summarize this discussion in the following proposition.

**Proposition 1** ***(Testability)*** *If the null hypothesis is weak, then SDPs have strong testable restrictions and BDPs do not. However, if the null hypothesis is strong, there are data that are behavioural-rationalizable and not standard-rationalizable, every outcome that is standard-rationalizable is also behavioral-rationalizable, and there are outcomes that are not behavioural-rationalizable.*

## 2.4 Reduced form Representation

The example of Section 2.2 aimed to illustrate in a simple way the positive and normative implications of the two key premises of our model: (a) preferences parameters change with actions and (b) people may not fully internalize this change. But how representative are these two premises of the existing literature? It can be shown that these two premises are enough to unify seemingly disconnected models in the literature, from situations where the preference parameter corresponds to the decision maker's current state (Tversky and Kahneman, 1991; Shalev, 2000), beliefs (Geanakoplos et al., 1989; Akerlof and Dickens, 1982), emotions (Bracha and Brown, 2008), expected consumption

(Koszegi, 2005; Koszegi and Rabin, 2006) or aspirations (Heifetz and Minelli, 2006), to models of adaptive preferences like Weizsäcker von (1971); Hammond (1976); Pollak (1978).

In this section, we review some of these models and show how our model can be obtained as a reduced form representation of them.

### 2.4.1 *Psychological games with a single active player*

Geanakoplos et al. (1989) (hereafter, GPS) study psychological games where the payoffs of each player depend not only on the actions chosen by all other players but also on what other players believe, on what she thinks they believe others believe and so on. Each player takes beliefs and actions of other players as given when choosing her own action. In equilibrium, beliefs are assumed to correspond to actions actually chosen. In the special case where there is a single active player, the payoffs of this single active player can depend on his own actions and the beliefs of other players over his own actions. Consider a two player psychological game. Player 1 is the active player with a set of pure actions $S$ and mixed actions $\Sigma = \Delta(S)$. A belief for player 2 is denoted by $\bar{b} \in \bar{B} = \Sigma$. The payoffs of player 1 over pure actions is given by a utility function $u : A \times \bar{B} \rightarrow \Re$ with $v(\sigma, b) = \sum_{s \in S} \sigma(s) u(s, b)$ being the corresponding payoffs over mixed actions. A psychological equilibrium is a pair $\left(\hat{\sigma}, \hat{b}\right) \in \Sigma \times \bar{B}$ s.t. (i) $\hat{b} = \hat{\sigma}$, (ii) for each $\sigma \in \Sigma$, $u\left(\hat{\sigma}, \hat{b}\right) \geq u\left(\sigma, \hat{b}\right)$. Clearly, by setting $A = P = \Sigma$ and $\pi$ as the identity map, a behavioral decision problem is a psychological game with one active player. GPS show that there are robust examples where the two sets $M$ and $E$ differ.

#### Example of self-image

Suppose Pat is concerned with what Jane will think about him. He can take a bold decision, which exposes him to the possibility of danger, or a timid, safe decision. Thus, his action space is $A = \{bold, timid\}$. Pat's payoffs not only depends on what he does but also on what he thinks Jane thinks about his character. In other words, Pat cares about what he thinks Jane thinks he will do. Suppose that Pat himself is a timid person, so he would prefer to think that Jane thinks he is timid rather than bold. But he also doesn't want to disappoint Jane, so if he thinks Jane expects him to be bold, he'd rather be bold than timid. Pat chooses bold with probability $b$ and timid with probability $1-b$. Let $q$ represent Jane's expectations of $b$ and $\tilde{q}$ represents Pat's expectations of $q$. The game and payoffs are described in the following matrix 2.11:

| | | | |
|---|---|---|---|
| | | $\tilde{q}$ | $1-\tilde{q}$ |
| | | *bold* | *timid* |
| $b$ | *bold* | $\underline{1}$ | 2 |
| $1-b$ | *timid* | 0 | $\underline{\mathbf{3}}$ |

(2.11)

The consistency requirement between actions and preference parameters of a behavioral equilibrium implies that we must have $b = q = \tilde{q}$. Beliefs must correspond to equilibrium play, yet they can still exercise a decisive influence on Pat's behavior. In this game, there are two welfare-ranked behavioral equilibria in pure strategies[14]. An optimal equilibrium in which $b = q = \tilde{q} = 0$ with payoff 3, and a sub-optimal equilibrium $b = q = \tilde{q} = 1$ with payoff 1.

It would be optimal for Pat to believe that Jane believes he is timid ($\tilde{q} = 0$), but if she does not believe so ($q = 1$), Pat will meet her beliefs and play bold ($b = 1$) reinforcing Jane's beliefs. In such case, Pat will end up doing something that he wouldn't do, was he able to internalize that playing timid would make Jane believe that he is indeed timid. If Pat reflected in that way, he would solve a standard decision problem in which $b = 0$ is the unique (Stackelberg) equilibrium. But in this behavioral decision problem, Pat is playing a la Nash in his intra-personal game against his self-2 who set the beliefs about what Jane believes. Self-2 "best reply" to Pat is to assign a probability equal to 1 to

[14] There is one equilibrium in mixed strategies, but we leave it aside for simplicity.

what Pat does.

### 2.4.2 *Loss aversion games with a single player*

Shalev (2000) considers a class of games where players have reference dependent utilities and the reference utility depends on the action profile chosen by all players. Shalev defines two notions of equilibrium, a myopic loss aversion equilibrium and a non-myopic loss aversion equilibrium. In either equilibrium notion, each player takes as given the actions of others when choosing her actions. In a myopic loss aversion equilibrium, a player also takes as given the reference utility when choosing her actions (even though changing her actions might change the reference utility). In a non-myopic loss aversion equilibrium, a player takes into account the feedback effect from her actions to the reference utility when choosing her actions. A single player version of Shalev's model has the player choosing a mixed action $\sigma \in \Sigma$ with payoffs $w(\sigma, r) = \sum_{s \in S} \sigma(s) v(u(s), r)$ where

$$v(u(s), r) = \begin{cases} u(s) & \text{if } u(s) \geq r \\ u(s) - \lambda (r - u(s)) & \text{if } u(s) < r \end{cases}, \tag{2.12}$$

$r$ is the reference utility and $u : S \to \Re$ is a standard utility function. A consistent reference point $r$ satisfies the equation $r = w(\sigma, r)$. Let $R(\sigma) = \{r \in \Re | r = w(\sigma, r)\}$. Shalev proves that $R(\sigma)$ is single valued and its values are contained in the closed interval $[\underline{r}, \overline{r}]$. Clearly, setting $A = \Sigma$, $P = [\underline{r}, \overline{r}]$ and $\pi(a) = R(\sigma)$, a behavioral decision problem is a myopic loss aversion decision problem while a non-myopic loss aversion decision problem corresponds to a standard decision problem. Shalev shows that in the static version of his model $M = E$ although the two sets differ in dynamic settings.

### 2.4.3 *Reference dependent consumption and personal equilibrium*

In Koszegi and Rabin (2006), a person's utility depends not only on her $K$-dimensional consumption bundle, $c$, but also on a reference bundle, $r$. She has an intrinsic "consumption utility" $m(c)$ that corresponds to the standard outcome-based utility. Overall utility is given by $u(c|r) = m(c) + n(c|r)$, where $n(c|r)$ is "gain-loss utility." In their chapter, both consumption utility and gain-loss utility are separable across dimensions, so that $m(c) = \sum_k m_k(c_k)$ and $n(c|r) = \sum_k n_k(c_k|r_k)$. They assume that $n_k(c_k|r_k) = \mu\left(m_k(c_k) - m_k(r_k)\right)$, where $\mu(.)$ satisfies the properties of Kahneman and Tversky's (1979) value function. Following Koszegi (2005) they define a personal equilibrium as a situation where the optimal $c$ computed conditional on forecasts of $r$ coincides with $r$. Clearly, by setting $A$ and $P$ to be the set of feasible consumption bundles and $\pi$ to be the identity map, a personal equilibrium can be represented by a behavioral decision problem[15]. Under the assumptions made in their paper, Koszegi and Rabin (2006) show that in deterministic settings $M = E$ while the two sets differ in stochastic settings

### 2.4.4 *Aspiration traps*

Heifetz and Minelli (2006) study a model of aspiration traps where an individual in period $t = 0$ makes a choice $e \in E'$, at a cost $c(e)$. For a given choice $e$, the decision problem of the individual at $t = 1$ is described by the tuple $G_e = \left(X, u_e, \bar{B}\right)$ where the strategy set of the individual is $X$, her payoff function is $u_e : X \times \bar{B} \to \Re$, and the utility of the individual depends on her attitude (beliefs, aspirations) $b \in \bar{B}$. When choosing a strategy $x(e, b)$ at $t = 1$ to maximize $u_e$, the individual takes as given both

[15] An analogous statement can be made for Koszegi and Rabin (2007), since the solution concepts they use (i.e. unacclimating personal equilibrium, UPE and preferred personal equilibrium, PPE) are examples of a "personal equilibrium" defined in Koszegi (2005). The major feature of these solution concepts is that the decision maker does not internalize the effect of her choice on her expectations (or reference point).

$b$ and $e$. However, given $e$, $b$ is determined by some preference formation mechanism $\beta : E' \to \bar{B}$. At $t = 0$, Heifetz and Minelli consider two modes of choice. When choice is "transparent", the individual would "see through" the preference formation mechanism. At $t = 0$, she would then choose $e$ to maximize $u_e(x(e, \beta(e))) - c(e)$. When the individual choice is "self-justifying", her choice of $e$ satisfies a no-regret condition

$$u_e(x(e, \beta(e))) - c(e) \geq u_e(x(e', \beta(e))) - c(e') \text{ for all } e' \in E'. \tag{2.13}$$

By setting $A = E'$, $P = \bar{B}$ and $\pi(a) = \beta(e)$, it is easily checked that a transparent choice problem corresponds to a standard decision problem while a self-justifying choice problem corresponds to a behavioral decision problem. Along the lines of example 1, they show that $M \subset E$.

### 2.4.5 *Adaptive Preferences*

Pollak (1978) defines the concepts of short run and long run demand functions associated with an adaptive preference mechanism. He models habit formation by assuming that a household's preference depend on its past consumption. Household's preference ordering in period $t$ is denoted by $\succeq_{c_{t-1}}$, where $c_t$ is the consumption vector for period $t$. The statement $c_t \succeq_{\bar{c}_{t-1}} c'_t$ means that the household finds $c_t$ at least as good as $c'_t$, given the consumption history $\bar{c}_{t-1}$. The *short-run* demand functions are denoted by $c_t = h(P_t, m_t, c_{t-1})$, where $P_t$ is the vector of prices in period $t$ and $m_t$ is the total expenditure in period $t$. The *long-run* demand functions $c = H(P, m)$ are defined to be the steady-state solution to the short-run demand functions: $H(P, m) = h[P, m, H(P, m)]$. What this older literature of the 70's including Weizsäcker von (1971); Hammond (1976); Pollak (1978) called long run preferences are equivalent to the preferences over actions given $p^*$, when $p^* = \pi(a^*)$.

## 2.5 Indistinguishability

How relevant is the distinction between a $BDP$ and a $SDP$? In this section, we state the conditions under which $BDP$ and $SDP$ are indistinguishable from each other. In those cases, $BDP$ and $SDP$ are outcome-equivalent and it is not worth making the distinction, although we show that this happens in very rare circumstances. In fact, in smooth settings, both decision problems are generically distinguishable. We then explore some peculiar outcomes of distinguishable decision problems.

### 2.5.1 Indistinguishability

A $BDP$ is indistinguishable from a $SDP$ if and only if $M = E$, i.e. the same choices are made in each model. Note that indistinguishability is, from a normative viewpoint, a compelling property. What matters for welfare purposes is the ranking of consistent decision states, which is the preference relation that a fully autonomous decision maker will use to make a decision. When $M = E$, the outcomes (consistent decision states) of a $SDP$ coincide with that of a $BDP$, and in that case, there is no reason to distinguish between normative and revealed preferences.

If $\pi(a) = \pi(a')$ for all $a, a' \in A$, a $BDP$ is, by construction, indistinguishable from a $SDP$[16]. So suppose the map $\pi$ has at least two distinct elements in its range. Next, consider the following conditions:

**C.1:** For $a, a' \in A$ such that $a \succeq_p a'$ for $p = \pi(a)$, $(a, p) \succeq (a', p')$ for each $p = \pi(a)$ and $p' = \pi(a')$;

**C.2:** For $(a, p), (a', p') \in \Omega$ such that $(a, p) \succeq (a', p')$, $(a, p) \succeq (a', p)$ for $p = \pi(a)$.

[16] In this case, all possible decision states $(a, p)$ are consistent and therefore, the procedure of choice (or the autonomy of the person) does not affect the outcome of the decision. Both, non-autonomous and fully autonomous decision makers will rank the outcomes in the same way and will choose the maximum outcome of the ranking.

**C.1** ensures that no behavioural equilibrium is sub-optimal, whereas **C.2** ensures that any standard equilibrium is also a behavioral equilibrium. Note that preferences in Example 1 violate **C.1** but satisfy **C.2** while the preferences in Example 2 violate both **C.1** and **C.2**. Shalev (2000) shows in Theorem 1 that in the static case his loss averse preferences satisfy both **C.1** and **C.2**. Koszegi and Rabin (2006) show that their reference dependent preferences also satisfy both **C.1** and **C.2**. GPS construct examples where, with one active player, both **C.1** and **C.2** are violated. Heifetz and Minelli (2006) construct examples where **C.1** is violated.

In the following proposition, we state that **C.1** and **C.2** are the necessary and sufficient conditions for indistinguishability.

**Proposition 2** ***(Indistinguishability)*** *Suppose that both $E$ and $M$ are non-empty. Then, (i) $E \subseteq M$ if and only if $(C1)$ holds. (ii) $M \subseteq E$ if and only if $(C2)$ holds.*

### 2.5.2 Smooth Decision Problems

To further understand the conditions under which indistinguishability occurs, it is convenient to look at smooth decision problems where decision outcomes are characterized by first-order conditions. We show that for the case of smooth decision problems, behavioral decisions are generically *distinguishable* from standard decisions.

A decision problem is smooth if (a) both $A$ and $P$ are convex, open sets in $\Re^k$ and $\Re^n$ respectively, (b) preferences over $A \times P$ are represented by a smooth, concave (component-wise) utility function $u : A \times P \to \Re$ and (c) the feedback map $\pi : A \to P$ is also smooth and concave.

A set of decision problems that satisfies the smoothness assumptions is *diverse* if and only if for each $(a, p) \in A \times P$ it contains the decision problem with utility function and feedback effect defined, in a neighborhood of $(a, p)$, by

$$u + \lambda p \tag{2.14}$$

and

$$\pi - \mu(a' - a) \tag{2.15}$$

for each $a'$ in a neighborhood of $a$ and for parameters $(\lambda, \mu)$ in a neighborhood of 0.

A property holds generically if and only if it holds for an open set of decision problems of full Lebesgue measure within the set of diverse smooth decision problems.

**Theorem 2** ***(Genericity)*** *For a diverse set of smooth decision problems, a standard decision problem is generically distinguishable from a behavioral decision problem.*

Eq. (3) shows in a simple quick way that $BDP$ and $SDP$ are indistinguishable only in isolated cases (e.g. when $\pi(a^*)$ or $u(a^*, p^*)$ are just constants).

Now that we know that making a distinction between $BDPs$ and $SDPs$ is a relevant route to take, we will explore some interesting peculiarities of distinguishable decision problems which pin down important policy implications. Our theoretical illustrations can be empirically complemented with Beshears et al. (2008), who describe situations in which revealed preferences deviate from normative preferences, or in our words, situations in which decision problems are distinguishable. They identify factors that increase the likelihood of having distinguishable decision problems, and discuss approaches to the identification of normative preferences when decision problems are distinguishable.

### 2.5.3 Distinguishable Decision Problems

We present a few examples that illustrate behavior that would be impossible to rationalize in a standard individual decision framework. In all these examples we assume, for simplicity, that $A$ and $P$ are finite sets and $\pi(a)$ is the identity map. The preferences of the decision maker are represented by an utility function $u : A \times P \rightarrow \Re$. We distinguish between pure and random behavioral decisions. Let $\beta(\hat{a}) = \arg\max_{a \in A} u(a, \hat{a})$. A **pure action behavioral equilibrium** is an action profile $a^*$ such that $a^* \in \beta(a^*)$. Let $\Delta(A)$ denote the set of probability distributions over the set of actions. A random strategy is $\sigma \in \Delta(A)$, where $\sigma(a)$ is the probability attached to action $a$. A random distribution over the set of psychological states is $\mu \in \Delta(A)$, where $\mu(\hat{a})$ is the probability attached to psychological state $\hat{a}$. A random decision state is a pair $(\sigma, \mu)$. Given a random decision state $(\sigma, \mu)$, the payoff to the decision maker is

$$w(\sigma, \mu) = \sum_{a \in A} \sum_{p \in P} \sigma(a)\mu(\hat{a})u(a, \hat{a}) \tag{2.16}$$

A consistent random decision state is a pair $(\sigma, \mu)$ where $\mu = \sigma$. A **random behavioral equilibrium** is a profile $\sigma^*$ such that $\sigma^* \in \arg\max_{\sigma \in \Delta(A)} w(\sigma, \sigma^*)$[17].

In each example, the decision problem is represented by a payoff table where rows are actions and columns are the psychological states. Under the assumptions made so far, consistent decision states are the diagonal of these payoff tables.

**Example 4** *A unique inefficient behavioral decision in dominant actions: addiction*

Consider the following payoff table:

| | $p_1$ | $p_2$ |
|---|---|---|
| $a_1$ | 1 | $-1$ |
| $a_2$ | 2 | 0 |

(2.17)

[17] I acknowledge feedback from the PhD examiners on the fact that this definition of random behavioural equilibrium assumes stochastic independence between actions and psychological states, invalidating the feedback effect. A perfect correlated equilibria might have been a better approach to undertake. In any case, a random behavioural equilibrium has been defined to introduce some examples as illustrations but the main results of this thesis are not affected by this definition.

We interpret these payoffs as an example of addiction where $a_2$ corresponds to *smoking* and $a_1$ corresponds to *not smoking* and $p_i$ to different health states of the individual ($p_2$ is less healthy than $p_1$). In this case, in a behavioral decision problem, the decision maker always chooses $a_2$ as $a_2$ is the dominant action for each value of $p$: if the individual takes her health state $p$ as given she always prefers to smoke. The unique behavioral decision outcome is $(a_2, p_2)$ with a payoff of 0. However, note that the consistent decision state $(a_1, p_1)$ with a payoff of 1 is the only element of $M$: once the individual takes the feedback from actions to health states into account, she always chooses not to smoke

**Example 5** *No pure action behavioral decision: the grass is always greener on the other side*

| | $p_1$ | $p_2$ |
|---|---|---|
| $a_1$ | 0 | 1 |
| $a_2$ | 1 | 0 |

(2.18)

We interpret these payoffs as an example of a situation where the individual makes a choice between two different lifestyle so that $p_i$ denotes a specific lifestyle and $a_i$ denotes the action that chooses location $p_i$. Starting from $p_1$, the decision-maker prefers $a_2$ to $a_1$ while starting from $p_2$, the decision-maker prefers $a_1$ to $a_2$: the individual always believes that the grass is greener on the other side. There is no behavioral decision in pure strategies. The decision-maker is, however, indifferent between both the two consistent decision-states $(a_1, p_1)$ and $(a_2, p_2)$.

This example demonstrates that, in general, $E$ may be empty even when $M$ isn't. However, given the discussion so far, a behavioral decision outcome can be interpreted as a Nash equilibrium of a two person game so that as long as $A$ and $Q$ are finite, a mixed strategy behavioral decision outcome always exists.

**Example 6** *Equilibrium in weakly dominated actions and domination by random actions*

| | $p_1$ | $p_2$ | $p_3$ |
|---|---|---|---|
| $a_1$ | 0 | 0 | 0 |
| $a_2$ | 0 | 1 | 2 |
| $a_3$ | 0 | 2 | 1 |

(2.19)

In this example, there are two behavioral equilibria, one in pure actions, $(a_1, p_1)$ and the other random, $(\frac{1}{2}a_2+\frac{1}{2}a_3,\, \frac{1}{2}p_2+\frac{1}{2}p_3)$. Note that in the pure action equilibrium $(a_1, p_1)$ the decision-maker is choosing a weakly dominated action and at the random equilibrium $(\frac{1}{2}a_2+\frac{1}{2}a_3,\, \frac{1}{2}p_2+\frac{1}{2}p_3)$, the decision-maker is strictly better off than at $(a_1, p_1)$. Note also that there is no pure action that (strictly) dominates $a_1$ although there are a continuum of random actions $qa_2 + (1-q)\,a_3$, $0 < q < 1$, that strictly dominates $a_1$.

**Example 7** *Multiple welfare ranked equilibria: aspirations*

| | $p_1$ | $p_2$ |
|---|---|---|
| $a_1$ | 1 | 0 |
| $a_2$ | 0 | 2 |

(2.20)

We interpret these payoffs as an example of an aspiration failure. Let $a_1$=*as undertaking an action that perpetuates the status quo* and $a_2$=*undertaking that changes the status quo*, with $p_2$ = *"high aspirations"* and $p_1$ = *"low aspirations"* being the consistent psychological states associated with $a_1$ and $a_2$ respectively. In this example, there are two strict behavioral decision outcomes $(a_1, p_1)$ and $(a_2, p_2)$. Note that the pure action equilibrium $(a_1, p_1)$ is dominated by the pure action equilibrium $(a_2, p_2)$. When decision-maker's aspirations are high, $(a_2, p_2) \succ (a_1, p_2)$, while when her aspirations are low, $(a_2, p_1) \succ (a_1, p_1)$. Thus, the behavioral decision outcome $(a_1, p_1)$ is an instance of an aspirations failure.

**Example 8** *More information may make the decision-maker worse-off*

Consider a decision problem with payoff relevant uncertainty, with two states of the world $\{\theta_1, \theta_2\}$ where the payoff tables are

$\theta_1 \rightarrow$ (2.21)

| | $p_1$ | $p_2$ | $p_3$ |
|---|---|---|---|
| $a_1$ | $-1$ | $0$ | $0$ |
| $a_2$ | $0$ | $3$ | $\frac{1}{2}$ |
| $a_3$ | $1$ | $4$ | $1$ |

$\theta_2 \rightarrow$ (2.22)

| | $p_1$ | $p_2$ | $p_3$ |
|---|---|---|---|
| $a_1$ | $1$ | $4$ | $1$ |
| $a_2$ | $\frac{1}{2}$ | $3$ | $0$ |
| $a_3$ | $0$ | $0$ | $-1$ |

Suppose, to begin with, the decision-maker has to choose before uncertainty is resolved. At the time when she makes the decision, the individual attaches a probability $\frac{1}{2}$ to $\theta_1$ and $\frac{1}{2}$ to $\theta_2$. In this case, expected payoff matrix is

(2.23)

| | $p_1$ | $p_2$ | $p_3$ |
|---|---|---|---|
| $a_1$ | $0$ | $2$ | $\frac{1}{2}$ |
| $a_2$ | $\frac{1}{4}$ | $3$ | $\frac{1}{4}$ |
| $a_3$ | $\frac{1}{2}$ | $2$ | $0$ |

It follows that the unique behavioral equilibrium is $(a_2, p_2)$ with expected payoff 3. Next, suppose that the decision-maker knows with probability one the true state of the world. Then, when the state of the world is $\theta_1$, $a_3$ strictly dominates all other actions and the unique behavioral equilibrium is $(a_3, p_3)$ with payoff 1 and when the state of the world is $\theta_2$, $a_1$ strictly dominates all other actions and the unique behavioral equilibrium is $(a_1, p_1)$ with payoff 1. Therefore, the decision-maker is worse-off with more information[18][19].

**Example 9** *Autonomy versus non-autonomy*

Consider the payoff table in matrix 2.18. In that example, if the decision maker took into account the feedback effect from actions to the utility parameter and maximized the induced utility function $v(.)$, $v(a_1) = v(a_2) = 0$. Therefore, a fully autonomous decision-maker who takes into account all the consequences of her actions would obtain a payoff of 0. However when the decision-maker doesn't take this feedback effect into account, we have already seen that there is a unique random outcome of the behavioral decision problem $(\frac{1}{2}a_1 + \frac{1}{2}a_2, \frac{1}{2}p_1 + \frac{1}{2}p_2)$ with an expected payoff of $\frac{1}{2} > 0$. On the face of it, it would seem that a non-autonomous decision-maker will be *better-off* than a fully autonomous decision-maker. But this interpretation isn't strictly true. In fact, if a fully autonomous decision maker is also allowed to choose mixed strategies in the payoff in matrix 2.18, she will also randomize $\{a_1, a_2\}$ by choosing the probability distribution $\{\frac{1}{2}, \frac{1}{2}\}$.

## 2.6 Coping Strategies

As defined in the introduction of this thesis, a complex decision is one in which (generally unobserved) payoff-relevant preference parameters are determined endogenously as part of the choice. All the examples and cases considered so far in this chapter assume that

[18] Note that in this example we are referring only to information that solves the uncertainty about exogenous states of the world. Our statement "the decision-maker is worse-off with more information" would not be right in the case in which additional information helps the decision-maker to control her own feedback effect, i.e. to gain autonomy.

[19] This result is consistent with Carrillo and Mariotti's (2000) results, although they use a dynamic model with time-inconsistent preferences.

the actions needed to affect the preference parameters are all in the action set. Take Section 2.2. example on reference-dependent preferences as an illustration of this point. The action set there is $A = \{a, q\}$ and the reference points are simply the expectation over actions. Since the person will adapt her expectations to her choices (though she does not anticipate so), whatever action she chooses from $A$ will determine her reference point. Same happens with all the models reviewed in Section 2.4. It may also be of interest to consider complex decisions in which the set of actions needed to change the relevant preference parameters of the decision problem is not accounted by the decision-maker as part of the problem. Here we present an example to illustrate this point.

Suppose that the action set is $A = \{a_1$=study, $a_2$=don't study$\}$ and the set of preference parameters is $P = \{p_1$=feel motivated to study, $p_2$= don't feel motivated to study$\}$. For the moment suppose that $p \in P$ is *exogenous.*

You have an exam tomorrow. There is no uncertainty and the options are very straightforward: if you study, you pass the exam; if you don't study, you don't pass it. Unfortunately, life is not as simple as this problem suggest. Your willingness to study depends also on your motivation. If you are lucky enough to feel motivated to study, you will study and get $u(a_1, p_1) = 3$. If you happen to feel unmotivated to study, you will not study and get $u(a_2, p_2) = 0$. In this case, since $p \in P$ is exogenous, there is nothing you can do to get $u(\cdot) = 3$, but pray for a higher motivation next time!

Suppose now that $p \in P$ is not purely exogenous, i.e. your motivation can be affected by your own choices. Suppose that you discover that after jogging in the park, you begin feeling motivated to study. So, if you are reflective enough, you will consider $A\prime = \{a_3 =$ jogging, $a_4 =$not jogging$\}$ to be part of your original problem of studying. Moreover, you should include the existence of a function $\pi$ that maps $A\prime \rightarrow P$, $a_3 \mapsto p_1$ and $a_4 \mapsto p_2$ as part of the decision problem. If you are a behavioural person, you will think that $p$ is exogenous to you. If you feel you are motivated to study, you will choose not to do jogging, sit-down, study and get $u(a_1, a_4, p_1) = 2$. However, by doing so you are making a mistake, i.e. you are disregarding the fact that your motivation is endogenous, and it comes only when you do jogging. Since you choose to study without jogging, this will trigger $p_2$ via $\pi$, and you end up in the worst equilibrium you could get, $(a_2, a_4, p_2)$, in which you are not motivated to study and you neither study nor do jogging. As long as you keep believing that your motivation is exogenous to you (belief sustained with imperfect reflection), you will be trapped in a self-defeating equilibrium.

This problem can be described in the following matrix:

| $u(a,p)$ | $p_1$ | $p_2$ |
|---|---|---|
| $(a_1, a_3)$ | $\mathbf{1}$ | $-1$ |
| $(a_2, a_3)$ | $0$ | $-1$ |
| $(a_1, a_4)$ | $\underline{2}$ | $-1$ |
| $(a_2, a_4)$ | $0$ | $\underline{0}$ |

(2.24)

The consistent decision sates are $\{(a_1, a_3, p_1), (a_2, a_3, p_1), (a_1, a_4, p_2), (a_2, a_4, p_2)\}$.There is one behavioural equilibrium, $(a_2, a_4, p_2)$ with utility $= 0$ and one standard equilibrium, $(a_1, a_3, p_1)$ with utility $= 1$.

Note that in this reinterpretation of the original problem, the net utility of passing the exam is $u(a_1, a_3, p_1) = 1$ instead of $u(a_1, p_1) = 3$ as it originally was. This is because we assume that reaching the right motivation to study has a cost, which is doing some jogging. Of course, life would be easier if we have the gift of always being cheaply motivated to do what we want. But, unfortunately, this is not always the case, and it would be beneficial to know how to deal with it.

Let's recapitulate and analyze what this story is telling us. The most important part is that the person is eventually able to escape on his own from this self-defeating equilibrium. In order to do so, she will have to re-interpret the problem in a different way. Namely, she must become aware that the decision problem is not anymore a simply choice of an $a \in A$ given some exogenous $p \in P$, but it is one in which she must choose a pair of actions $(a, a') \in A \times A'$, and an endogenous $p \in P$ with an associated map $\pi$. When she identifies that $p \in P$ is not exogenous as originally thought, she has made the first important step towards a change: realizing that there exists a map and she would make a mistake if she does not internalize the feedback effect. At this time she will know that she can do something to change his motivation, but still she may not know how. The second step is, then, to identify $A'$ and the map $\pi$. She needs to become aware that it is by jogging and not by doing other activity that brings her the right motivation to study. In other words, she needs to identify his coping strategy. Depending on the complexity of the map $\pi$, this learning exercise could be easier to grasp with therapy sessions. Alternatively, this learning outcome could eventually be achieved from observing someone similar to the decision maker who has already experimenting the map $\pi$.

Coping strategies are also mentioned in models of addiction and dynamic-inconsistent preferences. Dynamic-inconsistent sophisticated people understand their self-control problem and actively try to cope with it by choosing actions that reduce the likelihood of encountering cues that trigger binges. There is evidence of people purchasing costly pre-commitment devices like self-imposed deadlines (Ariely and Wertenbroch, 2002) or small packages of junk food, even when the unit price is lower for larger packages (Wertenbroch, 1998).

## 2.7 Psychological and Philosophical grounds for our Premises

Our framework relies on three key conceptual ideas. First, there is a feedback effect from actions to preference parameters that may not be fully internalized by the decision maker. Second, the individual's best interest is defined in the space of outcomes only when the feedback effect is internalized. Third, the individual always chooses what she judges best for her. In this section, we briefly review part of the literature in Social Psychology and the moral Philosophy that supports these conceptual ideas.

On the Social Psychology front, there is extensive work led by Albert Bandura who views human functioning as the product of a dynamic interplay of personal, behavioral, and environmental influence. Bandura points out that the way in which people interpret the results of their own behaviour informs and alters their environments and personal factors which, in turn, inform and alter subsequent behaviour through an "environmental feedback effect." He labelled this view "reciprocal determinism" (see Bandura, 1997, 2001). In line with Bandura's theory, there is a great deal of work favouring the hypothesis that the individual's actions may affect preference parameters. For example, Lazarus and Folkman (1984) argue that people are able to cope with stress, anger or anxiety, by changing their response to a situation (emotion-focused problem) or by changing the environment (problem-focused coping). Baron (2006, pg. 68) shows that emotions are partly under our control and argues that individuals can "induce or suppress emotions in themselves almost on cue." Baron argues that some people may even reshape their character, so that their emotional responses change.

On the philosophical front, the state of acting against one's better judgment has been studied since Plato and it has been labelled "Akrasia[20]". In the dialogue, Socrates sustains that "akrasia" is an illogical moral concept, claiming "No one goes willingly toward the bad" (358d). If a person examines a situation and decides to act in the way she determines to be best, she will actively pursue this action. In accordance to the normative principle advocated in this chapter, Socrates postulated that an all-things-considered assessment of the situation will bring full knowledge of a decision's outcome and worth linked to well-developed principles of the good. Davidson (2001), a contemporary American philosopher, argued that when people act in "akrasia" they

[20] In ancient Greek: Akrasia means "lacking command" (over oneself)

temporarily believe that the worse course of action is better, because they have not made an all-things-considered judgment, but only a judgment based on a subset of possible considerations.

The concept of personal autonomy has been subject of study specially in the literature of Philosophy (Friedman, 2003) and Psychology (Ryan and Deci, 2006). As Ryan and Deci (2006) point out, an act to be autonomous it must be endorsed by the self, fully identified with and owned. They also stress that "autonomy is not defined by the absence of external influences but rather by one's assent to such influences or inputs. Autonomy is thus not equivalent to independence." (pg. 1561). When autonomously functioning, people are more deeply engaged and productive, generating human capital and welfare (Woo, 1984). Ekstrom (2005) stress that autonomous acts proceed from one's core self, representing those preferences and values that are wholeheartedly endorsed. Dworkin (1988) maintains that people are autonomous only to the extent that their first order motives are endorsed at a higher order of reflection. As Ryan and Deci (2006) illustrate, a man who decides to "have another drink" would not be autonomous unless, in reflecting on this motive, he could fully endorse it. A lack of full endorsement would imply that the act is not autonomous. Moreover, Dworkin (1988) underscores that autonomy does not require behaving without or against constraints. For example, "although one might feel constrained in stopping for a school bus, if one assents to the value of traffic laws for ensure children's safety, one could willingly consent to the constraint and, in doing so, lose no autonomy."

As for the empirical part, self-determination theory (SDT) in Psychology provides a comprehensive picture of the importance of autonomy for well-being. Autonomy is considered a basic psychological need.

## 2.8 Conclusion

We offered a simple individual decision-making framework which is suitable for the study of general positive and normative implications of Behavioural Economics.

We made four contributions to the existing literature. First, we unified seemingly disconnected models, from more recent positive Behavioural Economics models to older models of adaptive preferences. Second, we proposed a new equilibrium existence result in pure actions without complete and transitive preferences. Third, we showed that behavioural decisions are falsifiable and that standard and behavioural decision problems have different testable implications. Fourth, we proved that in almost all the cases, behavioural and standard decisions are distinguishable from each other.

The chapter leaves some theoretical and empirical questions unanswered. We shall address some of these questions in the following chapters and identify others for future research in Chapter 7.

# Chapter 3

# Behavioural Decisions: Normative Analysis

## 3.1 Introduction

This chapter studies the normative implications of the framework proposed in this thesis.

Standard Normative Economics employs the revealed preference approach to extract welfare measures from behaviour. The *preferences revealed* from the individual's choices are assumed to be identical to the *normative preferences* representing the individual's true interest. People are assumed to choose what is best for them. Vast empirical evidence, however, has identified an array of situations in which individuals often do not do what is best for themselves establishing a wedge between normative and revealed preferences[1]. This situation raises a fundamental problem for economic welfare analysis: how can an appropriate criterion for analysing welfare be identified if the individual's

[1] Some of this evidence has been highlighted in the introduction of this thesis. For more evidence, see Loewenstein and Ubel (2008) who point out that in the "heat of the moment," people often take actions that they would not have intended to take and they soon come to regret. Koszegi and Rabin (2008) give examples of people making systematic mistakes (gambler's fallacy, projection bias, etc). Bernheim and Rangel (2005) record situations in which it is clear that people act against themselves: an anorexic refusal to eat; some people save less than what they would like; fail to take advantage of low interest loans available through life insurance policies; unsuccessfully attempt to quit smoking; maintain substantial balances on high-interest credit cards; etc.

choices fail to provide clear guidance on the individual's best interest?

In this chapter, we aim to contribute to this discussion by making use of the framework introduced in Chapter 1. Our approach is the following. As it is tradition in Economics, we argue that the person has "true or normative" preferences which she satisfies under perfect reflection. In Friedman's (2003) words "self-reflection is the process in which, roughly, a whole self takes a stance toward particular wants and values she finds herself to have" (p. 5). With perfect reflection the person fully characterizes the objects of choice, ranks them using his or her "true preferences" and chooses the object that maximizes this ranking. Perfect reflection ensures that "normative" and "revealed" preferences are equivalent concepts. However, as soon as we allow for imperfect reflection, the wedge between the two concepts may certainly arise challenging traditional public Economics. Indeed, Proposition 4 (pg. 56) of this chapter shows that in general, without perfect reflection and without further restrictions/information on the feedback effect, "revealed preferences" from choices do not represent "normative preferences." Nevertheless, we offer the necessary and sufficient conditions under which, regardless the degree of reflection, "revealed preferences" over actions are valid to construct welfare rankings. It turns out, as expected, that these conditions are very strong.

In the lens of our framework, a perfect reflective person must (a) acknowledge the existence of a feedback map (i.e. her preference parameter is not exogenous and she can do something to affect it), (b) identify her coping strategies (e.g. the set of actions that can be taken to affect the preferences parameters), (c) identify how such strategies affect her preference parameters (i.e. fully identification of the feedback map) and (d) implement her strategies efficiently. Observe that reflection implies autonomy. As argued by philosopher Friedman (2003, pg. 4) "autonomous behaviour is based on the deeper wants and commitments of the person, is partly caused by her reflections on and reaffirmations of them." To achieve autonomy a person must first somehow reflect on her wants and take up an evaluative stance with respect to them. Friedman (2003, pg.

4) makes very clear the endogeneity of preference parameters when she argues that "for choices and actions to be autonomous, the choosing and acting self as the particular self she is must play a role in determining them [...] When wants and desires lead to choice or action without having been self-reflectively endorsed by the person, the resulting choices and actions are not autonomous." By the same token, Elster (1983, pg. 22) defines autonomous preferences (or desires) as those "that have been deliberately chosen, acquired or modified – either by an act of will or by a process of character planning".

In the intra-self game-theoretic interpretation of our framework, being autonomous can be understood as playing a la "Stackelberg" and being non-autonomous as playing a la "Cournot." So, what matters for welfare analysis is how the person plays her intra-self game. If the person plays her game a la Stackelberg, then the revealed preferences from choices will be equivalent to her normative preferences, and therefore, can guide normative judgments.

We begin this chapter by relating our framework to choice theoretic models (Bernheim and Rangel, 2009, 2008; Rubinstein and Salant, 2008)[2] in which observed choices are determined by frames or ancillary conditions (Sections 3.2 and 3.3). The main concern of these models is whether such frames or ancillary conditions matter from a welfare viewpoint. There are at least two diametrically opposite views: (a) the choices of an autonomous decision maker should not depend on frames or ancillary conditions and therefore when frames and ancillary conditions do affect choices they could, in principle, matter for welfare, and (b) frames or ancillary conditions are exogenous to choices (if they appear endogenous this only because the decision problem isn't correctly specified) and are welfare irrelevant from a social planner's viewpoint[3].

---

[2]Hencerford BR and RS respectively.

[3]RS define a frame as "observable information that is irrelevant in the rational assessment of the alternatives, but nonetheless affects choice" (RS, 2008, abstract). For BR an ancillary condition is "an exogenous feature of the choice environment that may affect behavior, but is not taken as relevant to a social planner's evaluation" (BR, 2008, p. 4). Further quotations: "(...) if the individual's behavior appears to determine the ancillary condition endogenously, the decision problem has been defined incorrectly. For example, if he can choose to make his selection .... under one of two conditions, A or B, then it is inappropriate to describe A and B as endogenous ancillary conditions. Rather, the

If the view is (a), we argue that the choice mappings studied in BR and RS correspond to a decision problem where each frame or ancillary condition is consistent with any action chosen by the decision-maker. In a BDP, the decision-maker treats the frame or ancillary condition as given while in a SDP frames or ancillary conditions are considered to be characteristics of actions and are fully taken into account. In Proposition 5 (pg. 59) we show that in a world the welfare criterion proposed by BR is equivalent to the binary relation revealed from choices of a non-autonomous person. In Proposition 8 (pg. 61) we show that (*i*) the set of outcomes of a SDP is a subset of the set of outcomes of a BDP and (*ii*) the set of welfare optima derived solely from observed choice contains all the actions corresponding to the solution of a SDP. We conclude that the autonomy criterion can be applied to BR's approach as an evidence-based criteria for pruning elements of the welfare-relevant domain.

For a general feedback map from actions to preference parameters, we show that the weak welfare optima derived from observed choice may have an empty intersection with, or exclude elements of, the set of actions that correspond to the outcomes of a SDP.

If, on the other hand, both frames and ancillary conditions are truly exogenous to choices and welfare irrelevant for a social planner (case (b)), we argue that the choice mappings studied in BR and RS are generated in a family of distinct decision problems each indexed by a different preference parameter. In this case, the feedback from actions to preference parameters is a constant, the outcomes of BDP and SDP coincide and the set of welfare optima derived solely from observed choice is exactly the actions corresponding to the solution of a SDP.

Next, in Section 3.4, we assess the autonomy criterion in light of the other existing non-choice base normative criteria such as opportunity (Sugden, 2004) or individual happiness (Kahneman et al., 1997). We argue that the autonomy criterion encompass

correct ancillary condition describes the two-stage decision process." (BR, 2008 p. 34); "treating a condition of choice as a welfare-relevant characteristic of available objects would seem to defy common-sense...classifying it as an ancillary condition should be relatively uncontroversial" (p. 35).

the above mentioned criteria in the following sense: if the person is autonomous (a) her choices will reveal her true preferences, (b) more opportunities will be always good and (c) her subjective well-being may be a valid measure of her welfare. If the person is not autonomous, then any of these criteria may not be appropriate as a normative criterion.

In Section 3.5 we analyze some policy implications of our argument. Clearly, when the outcomes of both BDP and SDP coincide, there is no case for any sort of intervention by a social planner. In contrast, in scenarios where the set of outcomes of a SDP is contained in the set of outcomes of a BDP, we argue that the "libertarian paternalism" approach to policy interventions, advocated by Sunstein and Thaler (2003, 2008), that only seeks to alter the frames or ancillary conditions of individuals might work. We finish this chapter of normative implications arguing that policy interventions should, in general, aim to ensure that decision-makers internalize the feedback from actions to preference parameters.

Finally it is important to mention that although the positive models are fairly different, the most closely related approach to our work is proposed by Manzini and Mariotti (2009). They propose a particular decision-making procedure in which decision-makers categorize alternatives before choosing (CTC). Certainly, CTC decision-makers are not perfectly reflective, and it is left to see to what extent CTC decision problems differ from BDPs in terms of their axiomatic structure. In an important contribution for welfare analysis, Manzini and Mariotti (2009) show that by means of choice data alone, they can *uniquely* recover the underlying binary relation maximized in the post-categorization stage. In that way, the CTC procedure can pin down a valid welfare ranking from choices, under the assumption that people make decisions categorizing and then choosing.

## 3.2 Choice with Ancillary Conditions or Frames

Bernheim and Rangel (2009, 2008) (hereafter BR) and Rubinstein and Salant (2008) (hereafter RS) model choice problems where observed choices are determined by frames (RS) or ancillary conditions (BR). Examples of frames or ancillary conditions include the order in which candidates are listed on a ballot, default alternatives, salience of the alternative, deadline for making a choice (RS) or the point in time at which a choice is made, the manner in which information or alternatives are presented, the labeling of a particular option as the "status-quo," the salience of a default option, or exposure to an anchor (BR). For both BR and RS, frames or ancillary conditions are exogenous to the individual at the point when choices are made. Formally, in both papers, binary relations are constructed solely on observed choice and it is shown that such derived binary relations can be used to rank actions available to the decision-maker from a welfare viewpoint[4].

We assume that both $A$ and $P$ are non-empty finite sets containing at least two elements each.

Both BR and RS study generalize (or extend) choice problems $(A, p)$ where $p$ is a frame or an ancillary condition. An individual's choices[5] are described by a correspondence $c(A', p) \subseteq A'$ where $A' \in \tilde{\mathbf{A}}$ the set of all non-empty subsets of $A$ and further, that $c(A', p)$ is non-empty for all pairs $(A', p)$. Define BR's binary relation $\mathbb{P}^*$ as $a\mathbb{P}^*b$[6] iff for all admissible $(A', p)$ with $a, b \in A'$, $b \notin c(A', p)$. In BR's words $a$ is said to be strictly unambiguously chosen over $b$[7]. Define $a\mathbb{R}^*b$ iff $\sim a\mathbb{P}^*b$: there is some generalized choice problem where both $a$ and $b$ are present and $a$ is chosen. Define $a\mathbb{I}^*b$ iff $a\mathbb{R}^*b$ and $a\mathbb{R}^*b$: there is some generalized choice problem where both $a$ and $b$ are present and $a$ is chosen and some other generalized choice problem where both $a$ and $b$ are present and $b$ is chosen. BR show that $\mathbb{R}^*$ is necessarily complete: for any $a$ and $b$ the individual must necessarily choose either $a$ or $b$ from any $(\{a, b\}, p)$. Moreover, they also show that

[4]There are some technical differences between the two frameworks (see next section). Further, BR focus on the welfare implications of such framework while RS relate their framework to classical model of choice.

[5]RS study choice functions while BR allow for choice correspondences.

[6]For ease of exposition, since we allow for choice correspondences and focus on normative implications of choice, we follow BR although we note that RS also derive a preference relation similar to $\mathbb{P}^*$.

[7]In words, the statement "$a\mathbb{P}^*b$" means that whenever $a$ and $b$ are available, $b$ is never chosen.

none of the binary relations need be transitive although they do show that $\mathbb{P}^*$ is acyclic i.e. for any $a_1$,...,$a_K$, if $a_k \mathbb{P}^* a_{k+1}$ then $\sim a_K \mathbb{P}^* a_1$. BR, then, go on to make the following definition:

**Definition 1** ***(Weak welfare optimum, BR, 2008)*** *It is possible to strictly improve on a choice $a \in A$ if there is $b \in A$ such that $b\mathbb{P}^*a$. When a strict improvement is impossible, a is defined as a weak welfare optimum.*

BR show the following result that underpins their welfare analysis:

**Proposition 3** ***(Fact 1, BR, 2008)*** *If $a \in c(A', p)$ for some $(A', p)$, then a is a weak welfare optimum.*

Our framework suggests that what matters for welfare purposes is the ranking over consistent decision states, $\Omega$. The important question is whether revealed preferences over actions can be used to rank consistent decision states as well. Example *4* (Section 2.5.3, pg. 39) shows that this is not always the case. In that example, $a_2$ is always chosen and $a_1$ is never chosen. Therefore, $a_2$ is a strict (and hence, weak) welfare optimum as defined by BR. However, the decision state $(a_2, p_2)$ is dominated by $(a_1, p_1)$ and so the individual's revealed preferences over actions cannot be used to rank consistent decision states in this example, and it is this latter raking that matters for welfare assessments.

The following proposition states a necessary and sufficient condition for revealed preferences to rank consistent decision states.

**Proposition 4** ***(Behavioural implies Standard)*** *Let $a \in A$ be a weak welfare optimum. Then, any consistent decision state containing a, weakly welfare dominates any other decision state containing $a' \neq a$, $a' \in A$ if only if condition **C.1** holds.*

Recall that condition **C.1** was defined in Chapter 2 to prove indistinguishability. It is a necessary and sufficient condition to ensure that no behavioural equilibrium is sub-optimal, i.e **C.1** ensures $E \subseteq M$. Proposition 4 is telling us that even if we lived in a world with imperfect reflection and without further restriction or information on the feedback map, BR's binary welfare relation would yet be representative of the true preferences of the decision-maker, if an only if her true preferences satisfy condition **C.1**. The strength of this requirement will depend on the specific positive theory to which BR's approach is applied. For example, **C.1** is satisfied by the reference dependent preferences introduced by Koszegi and Rabin (2006) or by the static version of Shalev's preferences with loss aversion (Shalev, 2000). However, **C.1** is not always satisfied in some psychological games (Geanakoplos et al., 1989) or in models of aspirations traps (Heifetz and Minelli, 2006).

### 3.2.1 Characteristics of the object of choice

Both BR and RS make the point that, in practice, it is difficult to draw a distinction between characteristics of elements in $A$ and variables in $P$ which could also be viewed as characteristics of elements in $A$. In any actual decision problem studied in their papers, an individual takes the frame or ancillary condition as given when choosing an action. As already pointed out in the introduction, there are two mutually exclusive interpretations of choice with frames or ancillary conditions.

If the starting point is that the choices of a fully reflective decision-maker would never depend on frames or ancillary conditions, it is possible to relate the choice scenarios in BR and RS to those studied here by assuming that $\pi(a) = P$ for all $a \in A$ as each frame or ancillary condition is consistent with any action chosen by the decision-maker. Therefore, the outcomes of a SDP which correspond to the choices of a fully reflective decision-maker do not depend on frames or ancillary conditions as all characteristics of actions are taken into account. In this way, the fully reflective decision-maker chooses a pair $(a, p) \in A \times P$. In contrast, in a BDP the objects of choice are $a \in A$ taking as given $p \in P$ and therefore frames or ancillary conditions impact on choices and may matter from a welfare viewpoint.

The following two examples distinguish between the outcomes of a standard and behavioral decision problem behavioral decision problems when $\pi(a) = P$ for all $a \in A$.

**Example 10** ***(Default Options)*** *A possible interpretation of $p$ emphasized in both BR and RS is that of a label attached to objects of choice (such as "default options"). Consider $A = \{a, a'\}$ and $P = \{p =$"$a$ is the default option", $p' =$"$a'$is the default option"$\}$. In a standard decision problem, the individual will consider the label as a characteristic of the available objects and choose the optimal pair $(a, p) \in A \times P$. In a behavioral decision problem, however, she will take the label as given when making a choice without considering it as a characteristic pertaining to the object, and may choose $a$ over $a'$ at $p$ and $a'$ over $a$ at $p'$.*

**Example 11** ***(Dynamic inconsistency)*** *If $P = \{t = 1, t = 2, t = 3\}$ and $A = \{a_1, a_2, a_3\}$ where $a_t =$"complete task at $t$, do nothing at $t' \neq t$", $t = 1, 2, 3$. In a standard decision problem, the individual will choose both $(a_t, t) \in A \times P$, while in a behavioral decision-problem the individual will take $t$ as given so that, for example, at $t = 1$, $a_2$ will be chosen while at $t = 2$, $a_3$ will be chosen, thus being dynamically inconsistent.*

Alternatively, if both frames and ancillary conditions are truly exogenous to choices and welfare irrelevant for a social planner, it is possible to relate the choice scenarios in BR and RS to those studied here by assuming that for all $a \in A$, $\pi(a) = p_c$ for some fixed $p_c \in P$. In this case, each $p \in P$ indexes a distinct decision problem (with a different $\pi$) and in each of these distinct decision problems $M = E$.

## 3.3 Choice and Welfare

The framework studied in this thesis implies that welfare ranking should take place over consistent decision states using the preference relation $\succeq$. The approach of RS and BR, on the other hand, is based solely on choice data. In this section, we compare and contrast the normative implications of our decision model with choice with frames or ancillary conditions.

We begin by assuming that $\pi(a) = P$ for all $a \in P$ and we maintain this restriction on $\pi$ for the next three propositions because this is the way we understand BR's and RS's approach in the lens of our framework.

What matters for welfare for BR and RS is the binary relation $\mathbb{P}^*$ constructed solely from choices. In contrast, what matters for welfare in our model is the ranking of consistent decision states. The important question then is whether the ranking over actions using their binary relation $\mathbb{P}^*$ coincides with the fixed underlying true preference relation $\succsim$ over the set of consistent decision states.

Let's introduce the following notation. For any $A' \in \tilde{\mathbf{A}}$, let $\hat{c}(A') \subseteq A'$ be defined as the choice correspondence of a standard (autonomous) decision maker:

$$\hat{c}(A') = \{a \in A' : (a,p) \succeq (a',p') \, \forall a' \in A' \text{ and } p' \in \pi(a')\} \tag{3.1}$$

and let $\tilde{c}(A') \subseteq A'$ be defined as the choice correspondence of a behavioural (non-autonomous) decision maker:

$$\tilde{c}(A') = \{a \in A' : (a,p) \succeq (a',p) \, \forall a' \in A' \text{ and some } p \in \pi(a)\}. \tag{3.2}$$

Define the "*revealed standard*" binary relation $a\hat{\mathbb{P}}^*b$ iff for all admissible $(A',p)$ with $a,b \in A'$, $b \notin \hat{c}(A')$ and define the "*revealed behavioral*" binary relation $a\tilde{\mathbb{P}}^*b$ iff for all admissible $(A',p)$ with $a,b \in A'$, $b \notin \tilde{c}(A')$ with $\hat{\mathbb{R}}^*, \widehat{\mathbb{I}}^*$ and $\tilde{\mathbb{R}}^*, \widetilde{\mathbb{I}}^*$ defined analogously.

We begin by stating the following result as an immediate consequence of the analysis presented so far:

**Proposition 5** ***(Equivalence of*** $\mathbb{P}^*$ ***and*** $\tilde{\mathbb{P}}^*$***)*** *Suppose* $\pi(a) = P$ *for all* $a \in A$*. Then,*

$\tilde{c}(A') = \cup_{p \in P} c(A', p)$ *and* $\tilde{\mathbb{P}}^* = \mathbb{P}^*$.

Proposition 5 tells us that in a world in which every frame or ancillary condition should be considered as a characteristic of the object of choice, then the welfare criterion proposed by BR is equivalent to the binary relation revealed from choices of a non-autonomous person.

In what follows, we assume that the decision maker always solves a behavioral decision problem so that observed choice is described by $\tilde{c}(A')$ for each $A' \subseteq A$. Let $\hat{W}$ denote the set of weak welfare optima corresponding to $\hat{\mathbb{P}}^*$ and let $\tilde{W}$ denote the set of weak welfare optima corresponding to $\mathbb{P}^*$ and $\tilde{\mathbb{P}}^*$. The following proposition examines the link between $\hat{W}$ and $\tilde{W}$.

**Proposition 6** ***(Standard implies Behavioural)*** *Suppose* $\pi(a) = P$ *for all* $a \in A$. *Then,* $M \subseteq E$, $\hat{c}(A') \subseteq \tilde{c}(A')$ *and* $\hat{W} \subseteq \tilde{W}$.[8]

**Corollary 1** ***(E is non-empty)*** *From Proposition 6 as long as* $M$ *is non-empty, so is* $E$.

**Proposition 7** ***(M is non-empty)*** *Moreover, if the strict preference relation* $\succ$ *over* $A \times P$ *corresponding to* $\succeq$ *is acyclic,* $M$ *(and hence* $E$*) is non-empty.*

The importance of Proposition 6 is that in a world in which every frame or ancillary condition should be considered as a characteristic of the object of choice, the set of weak welfare optima derived solely from choice contains all the actions corresponding to outcomes of a SDP. Basically, in such a world, any standard equilibrium is also a behavioural equilibrium (i.e. **C.2** is satisfied). The problem here is that the reverse of this statement may not be true. There may be weak welfare optima derived from observed choices which do not correspond to outcomes of a SDP. The following example illustrates such case:

[8]Recall that $M$ is the set of all outcomes $(a, p)$ of a standard decision problem and $E$ is the set of all outcomes $(a, p)$ of a behavioral decision problem.

**Example 12** ***(Aspirations)*** *Modify the example 7 (Section 2.5.3, pg. 41) so that* $\pi(a) = P$ *for all* $a \in A$ *with no other changes. In this case,* $\hat{W} = \{a_1\} \subset \tilde{W} = \{a_1, a_2\}$. *Therefore,* $a_2 \in \tilde{W}$ *but* $a_2 \notin \hat{W}$.

One of the main criticisms of BR's normative criterion is that the set of individual weak welfare optima can become very large and even may get to have the same dimension of the set of alternatives. Proposition 6 identifies *autonomy* as a possible compelling criteria for pruning GCSs from the welfare-relevant domain (see BR, 2008, pg. 33).

Recall that so far we have restricted the feedback map to be the entire set of frames or ancillary conditions, following the interpretation that every frame or ancillary condition should be considered as a characteristic of the object of choice. In Proposition 8 we relax this assumption to examine the conditions under which the conclusion derived in Proposition 6 extends to the case when $\pi(a) \subset P$ for some $a \in A$. In this case, it is possible that the weak welfare optima derived from observed choices may have an empty intersection with actions corresponding to elements in $M$.

In Example 4 (Section 2.5.3, pg. 39), $\hat{W} = \{a_1\} \cap \tilde{W} = \{a_2\}$ is empty whereas in Example 5 (Section 2.5.3, pg. 40), $\hat{W} = \{a_1, a_2\}$ but $\tilde{W}$ is the empty set. In both examples, the weak welfare optima derived from observed choice have no connection with the actions corresponding to maximal consistent decision states. However, if both examples are modified so that $\pi(a) = P$ for all $a \in A$ with no other changes, then, in the now modified Example 4, $\hat{W} = \tilde{W} = \{a_2\}$ while in the modified Example 5, $\hat{W} = \tilde{W} = \{a_1, a_2\}$.

The following proposition states that a necessary and sufficient condition to obtain this conclusion for a general $\pi(.)$ is $C.2$.

**Proposition 8** ***(Standard implies Behavioural. The general case)*** *Suppose* $\pi(a) \subset P$ *for some* $a \in A$. *Then,* $\hat{W} \subseteq \tilde{W}$ *iff* $C.2$.

Finally, when for all $a \in A$, $\pi(a) = p_c$ for some fixed $p_c \in P$. In this case, each

$p \in P$ indexes a distinct decision problem and in each of distinct decision problems, $M = E$. For each $p \in P$, and any $A' \in \tilde{\mathbf{A}}$, let $\hat{c}(A', p) \subseteq A'$ be defined as the choice correspondence of a standard decision maker and let $\tilde{c}(A', p) \subseteq A'$ be defined as the choice correspondence of a behavioral decision maker. Clearly $c(A', p) = \hat{c}(A', p) = \tilde{c}(A', p)$ and the corresponding set of weak welfare optima all coincide.

## 3.4 Non-choice-based Welfare Criteria

As shown above, in general, preferences revealed from choices may not represent a clear guidance for making welfare assessments in our model, even if we applied BR's generalized criterion. Alternative criteria that base welfare on the individual's happiness or opportunity rather than choices avoid dealing with revealed preferences' problems. Below we briefly review these alternative criteria in light of our framework, and argue that they would not seem to provide unambiguous welfare guidance if we relaxed the assumption of perfect reflection.

### 3.4.1 Criterion based on Happiness or Experienced Utility

Kahneman et al. (1997) distinguish between *experienced utility*, which is a revival of the original hedonic conception of cardinal utility proposed by Bentham in 1789, and *decision utility*, which reflects the attractiveness of options as inferred from the individual's decisions. They claim that experienced utility is both measurable and empirically distinct from decision utility, and they propose to use the former as a relevant criterion for evaluating outcomes. Kahneman (2000) argues that experienced utility is best measured by moment-based methods that assess the experience of the present, such as self-reports of current well-being. The question that is generally asked to measure subjective well-being is: "how happy are you, overall?"

An important concern to explore in this context is whether the measures of "subjective well-being" of an autonomous individual are comparable to those measures of a non-autonomous individual. We conjecture that the answer is No. Here we present our argument that is only speculative, acknowledging that future research on this topic is highly important and needed.

Take Example 4 (pg. 39), with the pay-off matrix 2.17:

| | $p_1$ | $p_2$ |
|---|---|---|
| $a_1$ | 1 | $-1$ |
| $a_2$ | 2 | 0 |

Suppose the decision-maker is non-autonomous and chooses $a_2$. If she is asked how happy she is, she may well assess her level of happiness relative to what she thinks she can achieve given a fixed reference point $p_2$. Since $0 > -1$ she might report a high subjective level of wellbeing because she certainly believes she cannot do better! She is as good as she can be, given that she is solving the game *a la* Nash. Maybe, the level of happiness reported is equivalent to the level of happiness reported by an autonomous decision-maker who would compare the utility she gets, 1, with the utility she could have gotten with an alternative action, 0. In both cases, the difference of utility is $+1$, so both, the autonomous and the non-autonomous person would in principle report the same level of happiness, although clearly the non-autonomous person should be happier if she was autonomous. There might be an intermediate situation, in which the decision maker, for some reason, understands that she could eventually solve this game *a la* Stackelberg, but she cannot. In this case, we conjecture that she might report lower levels of happiness than if she had not realized about the endogeneity of her preference parameters. The reason being that now she would be able to compare the happiness she could get if she changed, 1, with what she is currently getting, 0[9]. These thoughts can be summarize in the following conjecture.

[9] This version of the model, however, does not allow a situation in which the decision maker is aware that she could play the game in a different way but she doesn't. This type of extension of the model will be considered for further research.

**Conjecture 1** ***(On subjective well-being)*** *Subjective well-being measures may provide ambiguous welfare guidance in the context of a BDP. If one still wanted to use such a measure, it should then be necessary to identify how autonomous the individual is at the moment to provide an answer to the subjective well-being question.*

### 3.4.2 Criterion based on Opportunities

Sugden (2004) proposes an "opportunity criterion" to guide welfare assessments. The normative principle behind his approach is that both the opportunity to choose between alternative options and the individual responsibility have moral value. Sugden's normative criterion is, therefore, not only based on the value that is attached to the "size and the richness" of the opportunity set, but also on the assumption that each individual "must take responsibility for how he uses his opportunities" (Sugden, 2004, pg. 1016). The conception of responsibility is crucial to provide value to "opportunity." A person is responsible if "at each moment in her life, she identifies with her own actions, past, present and future" (pg. 1018) whether or not her preferences change. As far as I understand, a responsible person can make mistakes, but recognizes them and internalizes their cost. So, it seems to me that some level of reflection is needed to be responsible, though responsibility does not necessarily imply autonomy. So, if a person can be responsible but "non-autonomous" then it is possible to show with an example that more opportunity may make the responsible (non-autonomous) person get worse.

**Example 13** ***On opportunities***

Consider first a situation where the payoff table is

| | $a_1$ | $a_2$ |
|---|---|---|
| $a_1$ | $-1$ | $0$ |
| $a_2$ | $0$ | $3$ |

(3.3)

In this case, the decision-maker has a unique efficient undominated action $a_2$ and there exists a corresponding outcome of the behavioral decision problem $(a_2, a_2)$ with payoff 3. Now, expand the set of choices so that the following payoff table represents the decision problem

| | $a_1$ | $a_2$ | $a_3$ |
|---|---|---|---|
| $a_1$ | $-1$ | 0 | 0 |
| $a_2$ | 0 | 3 | 1 |
| $a_3$ | 1 | 4 | 2 |

(3.4)

Note that $a_2$ continues to strictly dominate $a_1$ although now $a_3$ strictly dominates both $a_1$ and $a_2$. The unique behavioral equilibrium is $(a_3, a_3)$ with payoff $2 < 3$. This means that although the action set of the decision-maker has been expanded so that (a) the ranking of existing actions is unaffected and (b) the new action strictly dominates all existing actions, the individual is made worse-off. Note that all these features are robust to arbitrary but small perturbations in payoffs. This example illustrates our next remark:

**Remark 1** ***(On opportunities)*** *Larger action sets may make the decision-maker worse-off.*

Intuitively, what matters for the welfare of a non-autonomous decision maker is *how* she manages to take advantage of and use the opportunities she has at hand. We regard this remark as being close to the essence of Amartya Sen's capability approach. Sen's (1985) capability approach is a framework for the evaluation of individual welfare. Sen proposes to assess people's welfare in terms of their *functioning* and *capabilities*. A functioning is defined as an achievement of a person, i.e. what she manages to do (or be), whereas a capability reflects the various functionings an individual can *potentially* achieve. The space of functionings is the space of activities or state of being and the space of capabilities is the space of potential activities or states of being. Functionings measure *realized welfare*, whereas capabilities measure *potential or feasible welfare.* In that sense, the set of outcomes of a $BDP$ could be interpreted as the set of functionings, whereas the set of outcomes of a $SDP$ can be thought to be the set of capabilities. Given a Decision Problem $D$, an individual would be able to achieve her maximum potential welfare when $E \cap M \neq \varnothing$ and $E \subseteq M$. Therefore, we would get back to a result analogous to Proposition 4 (pg. 56), which would state that the individual will achieve her potentiality if and only if **C.1** holds (i.e. $E \subseteq M$). However, these are only conjectures, and the link between Sen's (1985) work and ours need further investigation.

The discussion in this section would suggest that the autonomy criterion refines some of the most widely used normative criteria, in the sense that the autonomy criterion implies the other criteria, but the reverse is not true.

**Remark 2** ***(Autonomy vis-a-vis other criteria)*** *If a choice has been made by a fully autonomous individual, then the preferences revealed from observed choices, the individual's happiness and her opportunities can be indistinguishably used as appropriate guidance for welfare judgements. However, the reverse of this statement is not true.*

Finally, it may be of interest to compute the (theoretical) *value of individual autonomy.* Take any decision problem $D$. Let $(\hat{a}, \hat{p})$ be the outcome of the $SDP$, $(a^*, p^*)$ the outcome of the $BDP$ and $\upsilon(a, p)$ the indirect utility function. The value of the individual's autonomy, $\alpha \in \Re_+$, is defined as:

$$\alpha = \upsilon\left(\hat{a}, \hat{p}\right) - \upsilon(a^*, p^*)$$

When the individual is fully autonomous, $\alpha = 0$. A social planner aiming to maximize individual welfare should choose a policy to minimize $\alpha$. The next section discusses some appropriate policies to achieve that aim.

## 3.5 Behavioral Public Policy

What are the policy implications of behavioral decision making, i.e. when people don't necessarily follow their own best interests?

The goal of any public policy is to maximize people's well-being. The route a social planner chooses to take in order to achieve that goal will depend on the social planner's presumption on *the way* individual chooses. In this section we (a) discuss the optimal extent of public policy, (b) claim that the existing public policy approaches do not necessarily maximize people's well-being and (c) introduce an alternative type of public policy based on empowerment.

### 3.5.1 Optimal extent of Public Policy

Before discussing what the appropriate policy interventions are, we wonder *when* a government intervention is necessary. In particular, is there an optimal extent of public intervention?

Take any Decision Scenario $D = (A, P, \pi)$. Let $v(a) = u(a, \pi(a))$. Now think of an individual maximizing:

$$Max_{\{a \in A\}} \mu v(a) + (1 - \mu) u(a, p) \tag{3.5}$$

where with some probability $\mu$ she takes the feedback effect from actions to preference parameters into account, and with some probability $(1 - \mu)$ she does not. Given this situation, let $\beta(p)$ denote the set of solutions to this maximization problem. An equilibrium of this problem is a decision state $(a^*, p^*)$ such that $(i)$ given $p^*$, $a^* \in \beta(p^*)$ and $(ii)$ $p^* = \pi(a^*)$.

The social planner's goal is to maximize $v(a)$ choosing an action $a \in A$. Suppose that the social planner has incomplete information about how reflective is the decision maker being, i.e. whether she is taking or not the feedback effect into account. In this context, the social planner maximizes:

$$Max_{\{a \in A\}} \mu' v(a) + (1 - \mu')\tilde{v}(a) \tag{3.6}$$

where with some probability $\mu'$ the social planner indeed chooses the $a \in A$ that maximizes the true interest or preference of the individual, and with some probability $(1 - \mu')$ the social planner uses a completely wrong set of preferences $\tilde{v}(\cdot) \neq v(\cdot)$.

**Remark 3** ***(Limits of intervention)*** *The extend of paternalism is limited by the trade-off between $\mu$ and $\mu'$. If the individual is internalizing the feedback effect with very high probability (high $\mu$) and the social planner has very few information about the individual (low $\mu'$), then there is no scope for intervention. On the other hand, if $\mu$ is low and $\mu'$ is high, then the social planner should intervene.*

How this trade-off can be identified is a relevant empirical question, although it is out of the scope of this chapter.

Note that, here, we have only considered the case of "hard" paternalism, in which the social planner chooses an action instead of the individual or, what is equivalent, forces the individual to choose a particular action[10]. However, the social planner can design *soft-libertarian* policies that allow the individual to internalize the feedback effect. We devote the next sub-section to explore the type of policy interventions implied by our framework.

### 3.5.2 Type of Policies

We begin this section going back to the original discussion we had the introduction about the endogenous preferences parameters as being part of the characteristics of the object

[10] Examples of paternalistic policies include banning narcotics, warnings on cigarettes, public health advertising, safety regulations such as the use of helmet or seatbelts, etc.

of choice or not. If for all $a \in A$, $\pi(a) = p_c$ for some fixed $p_c \in P$ and each $p \in P$ indexes a distinct decision problem with a different $\pi$, the binary relation derived from choice data alone $\mathbb{P}^*$ is the only one that is appropriate for constructing any welfare metric.

What are the policy implications of our model when $\pi(a) = P$ for all $a \in A$? As we have shown in Proposition 6 (pg. 60), in this case necessarily, $M \subseteq E$. With complete information about the individual preferences, if individuals choose $a$ taking $p$ as given, policy interventions that aim to shift the reference points of individuals (the $p's$) will ensure that individuals eventually choose actions that correspond to choices in $M$. So, our model can provide a set of conditions (complete information about preferences and $\pi(a) = P$ for all $a \in A$) that justifies the "soft paternalism" approach to policy interventions recently advocated by Sunstein and Thaler (2003, 2008).

However, even in this case, there is a different form of intervention, namely relaxing the internal constraint that the individual takes $p$ as given instead of choosing both $a$ and $p$, that would achieve the same objective but wouldn't rely on any information about the preferences of the individual.

In general, however, with incomplete information about individual preferences and feedback effects, direct policy intervention along the lines of "soft" paternalism (changing frames) or "hard" paternalism (making choices directly for the individual) could make matters worse.

One possible policy recommendation in scenarios with incomplete information about an individual's preferences is to directly act on the way in which a person internalizes the feedback effect from actions to frames[11].

To fix ideas, consider example *4*. In this example, if the individual doesn't take the feedback effect from actions to psychological states into account, she always chooses $a_2$ (*smoking*) over $a_1$ *(not smoking)*; however, the reverse would be true, if she took the feedback into account. Let $\alpha$, $0 \leq \alpha \leq 1$, denote the probability with which the

[11] An example of such policies could be psychotherapy sessions, projects aiming to foster people's emotional intelligence and empowerment, etc.

individual does take the feedback effect into account. A straightforward computation shows that as long as $\alpha > \frac{1}{2}$, the individual will choose $a_1$ over $a_2$: as long as the individual takes the feedback effect into account with a high enough probability, she will choose not to smoke.

Likewise, in example *7* (pg.41), a policy intervention that might work may be an "empowerment" policy, that would help the individual to become aware of her "internal constraints" and thus "gaining control over her own life[12]." We shall label these policies "*soft-libertarian*" which stands in between a fully libertarian and a libertarian paternalistic approach. A soft-libertarian policy intervention for efficiency purposes has the following features:

(a) It is only justifiable when

(i) condition **C.1** does not hold, i.e., choices cannot be used to guide welfare judgments,

(ii) the probability that the person is fully autonomous, $\mu$, is low,

(iii) the probability that the social planer knows $v(a)$, $\mu'$, is high.

(b) the intervention should not be coercive. On the contrary, it should aim to enhance the non-cognitive abilities needed to change the way a person solves her different intra-personal games: from Nash to Stackelberg. If the intra-personal game is one of aspirations traps, then a soft libertarian policy should affect the capacity to aspire. Empowering the person is one example of such policy. Stern et al. (2005) defines empowerment policies as those that help the person to "gain control over her own life." In the same line, Mullainathan (2006) argues that "good institutions also help to reduce problems that arises within a person." Likewise, Duflo (2006) claims that "what is needed is a theory of how poverty influences the decision making, not only by affecting the constraints, but also by changing the decision making process itself." If the intra-

[12] See for instance Appadurai (2004) on the "capacity to aspire" and Narayan (2002) or Stern et al. (2005) on "empowerment".

personal game is one of temptation, then a soft-libertarian policy should facilitate the capacity to exert self-control. Some examples of such soft-libertarian policies can be found in Bernheim and Rangel (2005). For example, behavioural therapies that teach cue-avoidance to addicts have shown to be successful. In our lens, by encouraging the adoption of new life-styles and the development of new interests, these therapies show the addict the way to become a Stackelberg leader. As pointed out by Bernheim and Rangel (2005), "these therapeutic strategies affect addict's choices without providing new information."

One conjecture is that this type of policies may have a more permanent impact on the individual well-being than the other existing "soft paternalistic" policies that propose changing the reference point exogenously, without helping the individual to do it herself. In fact, psychological studies show that autonomy support leads to greater program involvement, adherence and maintained change for behaviors such as smoking cessation, weight loss, glucose control and exercise (see Williams et al., 2005)

Finally, it is important to highlight that some standard policies that have always thought to be (at least weakly) welfare improving, may fail in our framework. For instance, as it was illustrated in examples 8 and 13 respectively, a policy that provides more information or more opportunities to a "non-autonomous" decision maker may make her worse-off. Or coming back to our example of aspirations, as Atkinson (1998) argues, creating jobs is not necessarily an effective policy to solve an aspiration failure: "ending social exclusion will depend on the nature of these new jobs. Do they restore a sense of control?"

## 3.6 Conclusion

We showed that revealed preferences cannot, in general, underpin welfare, though we offered conditions validating such a link. When choices fail to reveal true preferences, we argued that normative evaluations should be based on the degree of autonomy of the person, defined by the extent to which the person controls her own psychological states. A policy aiming at improving individual welfare should then look for developing people's abilities needed to make autonomous decisions. Examples of such abilities are capacities to aspire, to exert self-control or to correct maladaptive beliefs. Even though the concept of autonomy has been widely studied in Philosophy and Psychology, to my knowledge, this is the first time that the concept is brought up to a formal economic framework.

We acknowledge that, at the moment, our work has limitations on the empirical front. One of the main challenges concern the identification and measurement of autonomy. Chapter 7 discusses some ideas to address these concerns in future research.

# Chapter 4

# Behavioural Games: Theory and Applications

## 4.1 Introduction

This chapter extends the individual decision-making model proposed in Chapter 2 to a N-person strategic setting. Scholars in Psychology and Social Psychology study how preferences parameters, as defined in this thesis, are not only affected by own individual actions (i.e. $A_i \rightarrow P_i$) but also by others' actions and preferences parameters (i.e. $A_{-i} \times P_{-i} \rightarrow P_i$). For instance, an employee can consider that an action of her boss ($a_{-i} \in A_{-i}$) or beliefs ($p_{-i} \in P_{-i}$) are unfair to her and then feel mistreated ($p_i \in P_i$). A non-autonomous employee will take her emotional state as given, act as a passive "victim" of her own emotional forces and may choose to engage in acts of sabotage which could eventually lead to self-defeating outcomes. An autonomous employee, on the contrary, will do her best to manage his anger before the anger controls her[1].

As Chapter 1 distinguishes Behavioural from Standard Decisions, this chapter distinguishes "psycho-social" from "standard" normal-form games. A "standard" normal-

[1] See Mckinney and Newman (2002) for an example of students not fully internalizing the feedback effect.

form game is a normal-form game with endogenous preference parameters played by autonomous players. A "psycho-social" game, instead, is a normal-form game with endogenous preference parameters played by non-autonomous players.

Most of the interesting theoretical aspects of "psycho-social games" have been explored in the 1-person case introduced in the two previous chapters. Therefore, the focus of this chapter is mostly on applications. By providing a general framework to incorporate insights from Social Psychology[2] into Game theory, we can study the implications of different social motivations on people's behaviour. Examples of such social motivations are endogenous beliefs (Geanakoplos et al., 1989), endogenous social preferences (Rabin, 1993; Charness and Rabin, 2002; Charness and Dufwenberg, 2006) or endogenous reference-dependent preferences (Shalev, 2000). The chapter develops two new applications of psycho-social games introducing the idea of commitment and empowerment into game theory.

In addition, we obtain some novel theoretical results. We study two different ways of associating "psycho-social" games to "standard" normal-form games. One way is to take any normal-form game and *add* the endogenous feedback effect as an extra argument on players' utility function, thus *broadening* the preference domain. So the idea here is that besides the material payoffs of the original normal-form game, there is an extra argument (emotions, beliefs, aspirations, values) which players care about and can eventually affect. In such setting, we define the notion of a normal-form game being embedded in a "psycho-social" game. Then, we take any arbitrary "standard" normal-form game and compare its set of equilibria with the set of "psycho-social" equilibria of the psycho-social games that are embedded into this arbitrary "standard" normal-form game. With the same argument used in Chapter 2 to show generic distinguishability of BDP and SDP, we show that, typically, the two sets are distinct from each other, i.e. in general, people play "psycho-social" games instead of "standard" normal-form games. Thus, it shouldn't be surprising that most experimental evidence in normal-form games

[2] See for example Bandura (2001) or Forgas et al. (2003)

reports that people don't play Nash's predictions.

Another way of modelling the same idea of a feedback effect is by taking any "standard" normal-form game and *restricting* (instead of broadening) the preference domain in some consistent way. If the players compute their best response using the restricted utility, they will not internalize the externality they impose on themselves. In this case, we show that a strict Nash equilibrium of the original "standard" normal-form game is robust to any kind of imperfect reflection. Even assuming the highest level of imperfect reflection, if a "standard" normal-form game has a unique strict Nash equilibrium, then this equilibrium remains regardless how unreflective the players are: "no rationality at all is required to arrive at a Nash equilibrium." (Aumann, 1997, pg. 4). This is true when the game has a unique strict pure strategy Nash equilibrium. If the normal-form game has multiple or mixed strategy equilibria, then we show that some degree of "broad rationality" is needed to[3].

We then explore the relevance of the assumption of imperfect reflection in some existing games with endogenous preferences. We find different results depending on the game in question. For instance, in Rabin (1993), no matter how unreflective the players are, the predictions of the games do not change. The same is true in Shalev (2000). However, in other games (e.g. with guilt or commitment) the assumption of imperfect reflection changes the predictions of the game.

Finally, using similar arguments applied in Chapter 2, we provide two different existence results of a psycho-social equilibrium: one in pure strategies with incomplete preferences and strategic complementarity between actions and preference parameters, and the other relaxing the assumption of strategic complementarity and showing existence of a psycho-social equilibrium in mixed strategies.

The rest of the chapter is structured as follows. In Section 4.2 we introduce the model in a simple way using examples taken from the existing literature. Section 4.3

[3] See Elster (1983) for a distinction between broad and thin rationality.

provides the general model and the existence results. Section 4.4 shows applications on commitment and empowerment. Section 4.5 associates "psycho-social" games with "standard" normal-form games and provides some novel general results. Section 4.6 studies the implications of imperfect reflection in some existing games and Section 4.7 concludes.

## 4.2 A simple model with examples

Consider a simultaneous move game with two players $i = 1, 2$ whose payoff relevant variables are:

$i)$ the action set $A = A_1 \times A_2$ where $A_i = [\underline{a}_i, \bar{a}_i]$ is an interval of $\Re$.

$ii)$ the set of utility parameters (psycho-social states) $P = P_1 \times P_2$, where $P_i$ is a subset of some metric space. For the moment, we will keep it simple and assume that $P_i \subset \Re$.

The preferences of each player $i$ are represented by a utility function $v_i : A \times P_i \rightarrow \Re$. Each player $i$ solves the following maximization problem:

$$Max_{\{a_i \in A_i\}} v_i(a_i, a_{-i}, p_i), \quad \text{for } a_{-i} \in A_{-i} \text{ and } p_i \in P_i \tag{4.1}$$

This is the normal-form game that we all have seen in Micro101. In addition, as in a complex decision problem (see pg. 25) suppose there is a feedback effect from the vector $(a_i, a_{-i}, p_{-i})$ to $p_i$ represented by the map $\pi_i : A \times P_{-i} \rightarrow P_i$. This map represents the idea that players can affect their own and others psychological states. If we assume that each player $i$ reflects perfectly, when they consider a deviation from $a_i$ to $a_i'$, they will anticipate the appropriate change in their and others psychosocial states consistent with $a_i'$, and solves the following optimization problem:

$$Max_{\{a_i \in A_i\}} u_i(a_i, a_{-i}), \text{ for } a_{-i} \in A_{-i} \tag{4.2}$$

where $u_i(a_i, a_{-i}) = v_i(a_i, \pi_i(a_i, a_{-i}, p_{-i}))$. We call this game, a "standard" normal-form game, and its solution concept is analogous to a Nash Equilibrium.

In our definition of a "psycho-social game", we shall assume that players do not take into account the consequences of their actions on their own and other's psychosocial states. When evaluating a deviation, the player will not change her psychosocial state but will take it as given.

In Chapter 1 we showed that Geanakoplos et al.'s (1989) (hereafter GPS) psychological games can be reduced to a BDP (see Section 2.4, p.g. 31). It is straight forward to show that "psycho-social" games generalize psychological games. Both games consider a preference parameter that is endogenously determined within the interaction, and in both games the parameter is taken as given at the moment to compute the best response. However, in GPS' framework preference parameters are only beliefs, whereas in our framework, preference parameters are not restricited to beliefs, they can be any possible endogenous states. This feature generalizes psychological games and allows to study endogenous states that are not necessarily belief-dependent (e.g. identity, aspiration, a good, the environment, etc.). Moreover, even if we wanted to work in the space of beliefs, psycho-social games are more general than psychological games because they also allow for 0-order beliefs (i.e. self-beliefs). These beliefs can be interpreted either as moral values (i.e. player's own beliefs about what she should play) or as player's self-confidence (i.e. player's own beliefs about she can play)[4]. We show some examples using 0-order beliefs in next section. But for the moment, we continue our description of other existing model that is generalized by psycho-social games.

[4]See Bandura (1997) or Bandura (2000) for references about how self-confidence affects subjective well-being.

### 4.2.1 Social Preferences

In our understanding, there are two existing ways to incorporate social preferences into strategic models: an *exogenous* "distributional" approach and an *endogenous* "reciprocity-guilt" approach. The former broadens players' preferences with an exogenous social preference parameter which captures how people care about others[5]. The latter, in contrast, allows social preference parameters to be determined endogenously within the strategic interaction. Rabin (1993) was the first paper who proposed this second approach by introducing endogenous reciprocity motives into players' preferences. A decade later, Charness and Rabin (2002) merged endogenous reciprocity models with exogenous distributional models. Recently, Charness and Dufwenberg (2006) introduced *guilt* as another endogenous motivation for social preferences.

"Psycho-social" games encompass these three models and offer a higher degree of freedom to capture other class of social preferences that cannot be captured in the existing literature. In what follows, we take Charness and Rabin (2002)[6] general model with $n-$players and show that it is a special case of a psycho-social game[7]. In next section we present a novel application of a psycho-social game in which there is an interaction between two players who care about each other but for different reasons. One has a moral commitment; the other feels sympathy.

Let $A_i$ be player $i's$ pure strategies and $A_{-i} = \prod_{j \neq i} A_j$ be the set of pure strategies of all players but player $i \in N$. The material payoffs are determined by the action profile $a \equiv (a_i, a_{-i})$ where $u_i(a)$ represents Player $i's$ payoffs given action profile $a \in A$. Let $p \equiv (p_i, p_{-i})$ be a *demerit* profile, with $p_i \in [0, 1]$ being a measure of the beliefs that all players but $i$ have about how much player $i$ deserves[8] and $p_{-i}$ being a vector $(p_1, ..., p_j, ..., p_n)$ for all $j \neq i$ representing player $i$'s disposition towards all the other players. The higher the value of $p_i$ the less player $i$ deserves. Given profiles $a$, $p_i$ and a

[5] See for example Fehr and Schmidt (1999).

[6] Henceforth C&R

[7] We will work in pure strategies for expositional purposes only.

[8] $p_i$ (resp. $p_{-i}$) here is $d_i$ (resp. $d_{-i}$) in Charness and Rabin's (2002) paper.

set of parameters $\kappa = (\lambda, \delta, b, k, f)$,player $i's$ preferences are defined as follows:

$$v_i(a, p_{-i}) \equiv (1-\lambda_i)u_i(a) + \lambda_i \left[ \begin{array}{c} \delta \left[\min \left\{ u_i(a), \min_{j\neq i} \left\{ u_j(a) + bp_j \right\} \right\}\right] + \\ (1-\delta) \left[ u_i(a) + \sum_{j\neq i} \max \left\{ 1 - kp_j, 0 \right\} u_j(a) \right] - \\ f \sum_{j\neq i} p_j u_j(a) \end{array} \right] \tag{4.3}$$

where $\lambda_i \in [0, 1]$ measures how much player $i$ *cares about pursuing the social welfare* versus his own self-interest; $\delta_i \in (0, 1)$ measures the degree of concern for helping the worst-off person versus maximizing the total surplus and $b$, $k$ and $f$ are nonnegative parameters.

Given $a_{-i}$, $p_{-i}$ and $\kappa$, the set of player $i's$ actions that maximize her utility is:

$$A_i^*(a_{-i}, p_{-i}; \kappa) \equiv \left\{ a_i^* \in A_i | \, a_i^* \in \arg\max v_i(a_i, a_{-i}, p_{-i}; \kappa) \right\} \tag{4.4}$$

So far the only endogenous variable is $a_i$ for all $i \in N$ and then, this is just a model with an extended utility that incorporates exogenous distributional concerns and exogenous concerns for reciprocity. If $p_i = 0$ for all $i \in N$, then this model becomes a simple model without psycho-social states. However, C&R endogenizes $p_i$ and $\lambda_i$. It is assumed that the way player $i$ cares about the welfare of others (i.e. $\lambda_i$) depends on the action profile and on how much player $i$ thinks the others deserve. Formally, $\lambda_i(a, p_{-i})$ is assumed to be an upper hemi-continuous and convex-valued correspondence from $(a, p_{-i})$ into the set $[0, 1]$ such that $\lambda_i(a, p_{-i}) \approx \{ \lambda | \, a_i^* \in A_i^*(a_{-i}, p_{-i}; \lambda) \}$. In turn, the function $\lambda_i(a, p_{-i})$ is a measure of *how appropriate other players feel player* $i$ *behaves* when they determine how to reciprocate.

C&R derive demerit profiles from these functions and assume that other players compare each $\lambda_i(a, p_i)$ with some exogenous selflessness standard $\hat{\lambda}$ - the weight they feel a *decent* person should put on social welfare. The equilibrium of this game is defined as follows:

The strategy profile $a^*$ is a *reciprocal-fairness equilibrium* if for a given parameter profile $\kappa = (\hat{\lambda}, \delta, b, k, f)$, the following conditions hold for all $i \in N$:

$i)$ given $a^*_{-i}, p^*_i$ and $p^*_{-i}$, $a^*_i \in \arg\max v_i(a_i, a^*_{-i}, p^*_i, p^*_{-i})$,

$ii)$ given $\lambda^*_i$, $p^*_i \in \arg\max[\hat{\lambda} - \lambda^*_i, 0]$

$iii)$ given $a^*_{-i}, a^*_i$ and $p^*_{-i}$, , $\lambda^*_i \in \lambda_i(a^*_i, a^*_{-i}, p^*_{-i})$

A reciprocal-fairness game is a special case of a psycho-social game in which the general map $\pi_i : A \times P \to \Re$ is a particular composite correspondence $(\lambda_i \circ p_i)(\kappa)$ that assigns a value from $\max[\hat{\lambda} - \lambda_i(a, p_{-i}), 0]$ to the interval $[0, 1]$. It is important to point out that when player $i$ chooses an optimal $a_i$, besides taking as given $a_{-i}$ and $p_{-i}$, she **takes as given** the endogenous $p_i(\lambda_i(a, p_{-i}))$ - i.e. how much other players think she deserves -, which in turn depends on her optimal action via the correspondence $\lambda_i$. Again, there is a feedback effect from player $i's$ action into her preference parameter that it is not internalized when choosing a best response. Thus a reciprocal-fairness game is a special case of a psycho-social game, *all reciprocal-fairness equilibria* are psycho-social equilibria and, by transitivity, all fairness equilibria (Rabin, 1993) are psycho-social equilibria too. In Rabin's (1993) model $p_i = (b_j, c_i)$.

C&R don't study the case in which a player may care about the others to avoid feeling *guilt*. However, guilt motivations can be represented by means of second order beliefs $p_i = c_i$, i.e. player $i's$ beliefs about what player $j$ believes player $i$ will do. A guilt averse person will try to minimize the *guilt* caused by not conforming to the expectations of others. For example, if a guilt averse person playing a public good game believes that the other players believe she will cooperate, then she may consider cooperating.

We are aware of the fact that *reciprocal-fairness* equilibria and *guilt* equilibria are not only "psycho-social" equilibria but also "psychological" equilibria (Geanakoplos et al., 1989). In Section 2.4 (pg. 33) we have shown that Loss-Aversion games (Shalev, 2000) can be reduced to BDP, and as such, they are generalized by psycho-social games.

However, loss-aversion games are not psychological games.

## 4.3 Psycho-social Games

### 4.3.1 The general model

Formally, the game is structured as follows. There is a finite set $N = \{1, ..., n\}$ of players (indexed by $i$) and for each player a finite set of (pure) actions $A_i$. In addition, each player is characterized by a preference parameter (or psycho-social state) $p_i \in P_i$, where $P_i$ is a corresponding subset of some metric space. Let $A = \prod_{i \in N} A_i$, $A_{-i} = \prod_{j \in N \setminus \{i\}} A_j$ and $P = \prod_{i \in N} P_i$, $P_{-i} = \prod_{j \in N \setminus \{i\}} P_j$. A generic element of $A_i$ (resp. $A$) is denoted by $a_i$ (resp. $a$) and a generic element of $P_i$ (resp. $P$) is denoted by $p_i$ (resp. $p$). It is assumed that $A \subset \Re^{nm_i}$ and $P \subset \Re^{n\rho_i}$ are subsets of a finite dimensional Euclidian space[9]. For the purposes of this chapter we shall consider only pure strategies.

A *consistent (pure) psycho-social state* for player $i$ is a $\tilde{p}_i \in P_i$ such that $\tilde{p}_i = \pi_i(a_i, a_{-i}, p_{-i})$. The set of (pure) consistent psycho-social states is

$$\tilde{P}_i = \left\{ \begin{array}{c} \tilde{p}_i |\, p_i = \pi_i(a_i, a_{-i}, p_{-i}), \\ \text{for all } i \in N,\ a \in A \text{ and } p_{-i} \in P_{-i} \end{array} \right\}.$$

For the purpose of equilibrium existence, we will require $\pi_i(a_i, a_{-i}, p_{-i})$ to be non empty and close relative to $P$ for each $a_{-i}$ and $p_{-i}$.

Player *$i's$ utility function* $v_i : A \times P_i \rightarrow \Re$ depends on the outcomes (as in the standard literature) and *also* her and other players preference parameters (or psycho-social states)[10]. We assume that player $i$ seeks to maximize $v_i$ given $p$ and $a_{-i}$.

[9] Note that if we wanted to include GPS psychological games as a particular case of a psycho-social game, we should consider the set $P_i$ to be a subset of a Polish space. However, this would invalidate our proof for existence in pure strategies and so, we rather assume that $P_i$ is a subset of a finite dimensional Euclidean Space.

[10] For expositional purposes, we assume in this section that preferences can be represented by a utility function. However, for our existence proof, we will not require preferences to be able to be represented by a utility function, i.e. complete and transitive.

**Definition 2** ***Game****: A normal-form game with endogenous preference parameters $\psi = (A_i, P_i; v_i, \pi_i)$ consists on an action set $A_i$,a set of utility parameters $P_i$, an utility function $v_i : A \times P_i \to \Re$, and a map $\pi_i : A \times P_{-i} \to P_i$ for each player $i$.*

**Definition 3** ***Equilibrium****: A (pure) psycho-social equilibrium of $\psi$ is defined by a pair $(a^*, p^*) \in A \times P$ if, for each $i$, $(i)$ given $p_i^*$, $p_{-i}^*$, $a_{-i}^*$, $a_i^* \in \arg\max v_i(a_i, a_{-i}^*, \tilde{p}_i^*)$ and $(ii)$ given $a_i^*$, $p_{-i}^*$ and $a_{-i}^*$, $\tilde{p}_i^* \in \pi_i(a_i^*, a_{-i}^*, p_{-i}^*)$*

### 4.3.2 Existence in pure strategies

In this section, we examine the conditions under which a psycho-social equilibrium in pure strategies exists. To be coherent with the model we have in mind, we allow for non-convex and incomplete set of preferences due to two main reasons. Firstly, since player's preferences depend on their psycho-social states which is a reference parameter, their preferences sets may not be convex. Secondly, in a context in which psychological states play a role, preferences may be more likely to be incomplete: it is more likely that I have a complete preference relation between pears, bananas and apples than between pears when I am stressed, bananas when I am sad and apples when I am anxious. For these reasons, we examine the conditions for equilibrium existence allowing for incomplete and non-convex preferences. Due to non-convexities we cannot apply Kakutani's fix-point theorem. We apply Tarski's fix point theorem instead. The proof is an extension of the existence proof introduced in chapter 2 to a strategic interaction setting. We assume that the set $A$ and $P$ are compact lattices, that the preference relation is acyclic and has an open lower section, that psycho-social states of each player are increasing in her actions, quasi-supermodularity, single-crossing property and monotone closure.

**Theorem 3** *Under the assumptions mentioned above, there exists a psycho-social equilibrium.*

### 4.3.3 Existence of a mixed psycho-social equilibrium

When we proved the existence of a psycho-social equilibrium in pure strategies, we paid the cost of AS-4 to AS-7. These assumptions are appropriate for some applications such as motivation and aspiration failures, but they do not fit other motivations (e.g. Example 2.18). Of course, with those assumptions we gained in other aspects. For example, we did not need $A$ and $P$ to be convex sets and preferences to be complete and transitive. In this section, we relax the assumptions of single-crossing, and we look for the existence of a mixed psycho-social equilibrium. As it was the case for the existence of an equilibrium in pure strategies, we will not require preferences to have an expected utility representation. Further, we will assume that preferences have open lower sections, which is a weaker continuity assumption than the standard assumption of continuity proposed in the literature but it is still stronger than (AS-2) made in Theorem 3.[11]

**Theorem 4** *Under assumptions (AS-1') to (AS-4') (see Appendix), a mixed strategy psycho-social equilibrium exists.*

## 4.4 Applications

### 4.4.1 Commitment and Sympathy

In "Rational Fools: A Critique of the Behavioral Foundations of Economic Theory," Sen (1977) discusses the view held in traditional Economics that "every agent is actuated only by self-interest." In his paper, Sen distinguishes two separate concepts: (i) *sympathy* and (ii) *commitment*. He argues that sympathy "corresponds to the case in which the concern for others directly affects one's own welfare. If the knowledge of torture of others makes

[11] In general it is assumed in the literature that preferences have open graphs (see Shafer and Sonnenschein (1975)).

you sick, it is a case of sympathy; if it does not make you feel personally worse off, but you think it is wrong and you are ready to do something to stop it, it is a case of commitment (p. 326)." Behaviour based on sympathy is in an important sense egoistic whereas the action based on commitment is non-egoistic. Most influential existing models on social preferences do not capture this distinction. Moreover, if they wanted to incorporate *individual commitment*, they would need to move from a psychological game to a psycho-social game setting. We introduce *moral commitment* into a game as *0-order beliefs* affecting payoffs. For some reason, players have their *own beliefs* about what is right and what is wrong, and they enjoy when they choose an action that is consistent with those beliefs. Let's consider a discrete version of the dictator game. There are two players. Player $i$ is the dictator who chooses either to keep (K) or give (G) to player $j$ his experimental endowment (say £1). Player $j$, the recipient, has no real choice—she has to simply accept the dictator's decision. Now assume that the dictator has a moral commitment with player $j$, and his utility depends purely on conforming to her (endogenous) commitments. Let $\tilde{p}_i \in [0,1]$ be player $i's$ beliefs about her own actions and $p_i$ be the probability attached to playing G.

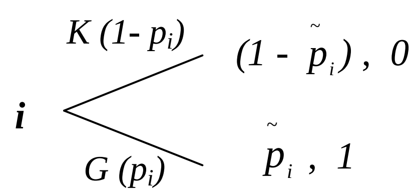


Figure 4.1: Commitment

This game has three psycho-social equilibrium: $p_i = \tilde{p}_i = 0$, $p_i = \tilde{p}_i = 1$ and $p_i = \tilde{p}_i = \frac{1}{2}$, with final payoffs being $1, 1$ and $\frac{1}{2}$ respectively.

When players have different motivations for social preferences, there is a possibility of misperception of players' kindness. A player might think that she is being kind, when in fact, the others interpret she is being unkind. The way in which player $i$ interprets others' intentions will depend on player $i's$ moral frame. We shall illustrate this point with the following example proposed by Sen, about two boys who find two apples, one large and one small.

> Boy A tells boy B, "You choose." B immediately picks the larger apple. A is upset and permits himself the remark that this was grossly unfair. "Why?" asks B. "Which one would you have chosen, if you were to choose rather than me?" "The smaller one, of course," A replies. B is now triumphant: "Then what are you complaining about? That's the one you've got!" B certainly wins this round of the argument, but in fact A would have lost nothing from B's choice had his own hypothetical choice of the smaller apple been based on sympathy as opposed to commitment. A's anger indicates that this was probably not the case (p. 328-9)."

This story can be modeled as a two-person game with only one active player chosen at random by nature with the following payoffs:

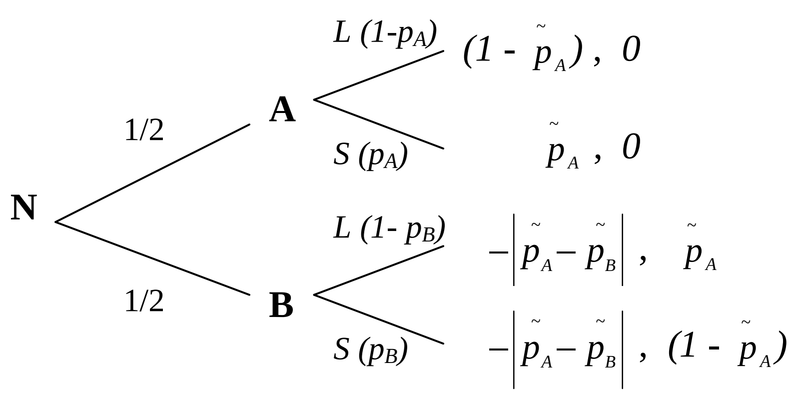


Figure 4.2: Commitment and Sympathy

$p_A$ (respectively $p_B$) represents the probability with which player $A$ (respectively $B$) chooses to keep the small apple $S$. For expositional purposes, we shall assume that the material payoff derived from the large apple ($u_i(L) = 0$) is the same than that derived from the small one ($u_i(S) = 0$). In the standard (trivial) version of this game with just material payoffs, any possible strategy profile is an equilibrium profile.

Suppose now that Player $A$ has *moral commitments* whereas player $B$ just *sympathizes* with $A$ and she cares about being fair with her.

For some reason, player $A$ has her *own beliefs* about what is right and what is wrong, and she is committed to choose actions consistent with her own beliefs. Moreover, she feels pleasure when $B$ chooses an action that matches her (player $A$) own beliefs.

Player $B$, on the other hand, likes to be fair with $A$ in her particular way. We introduce fairness through *1st-order beliefs*. Player $B$ enjoys giving player $A$ what player $A$ would choose to keep for her if she was to choose rather than player $B$[12].

Let's see the payoffs of this game. If player $A$ believes that keeping the small apple for her (i.e. playing $S$) is the right thing to do, then $p_A = 1$. So if she had to make a choice, given her *values*, she would choose $S$. If player $B$ is the one called to choose, $A's$ payoff is higher the closer is $B's$ action to $A's$ values. In other words, $A$ prefers $B$ choosing $p_B = 1$ rather than $p_B < 1$, since $p_B = p_A$ means that both players have the same values and player $A$ enjoys that. Now looking at player $B's$ payoffs, she does not feel either negative or positive emotion regardless what player $A$ chooses, since she just care about being fair with player $A$. If she had to choose, she would choose to keep the large apple, $L$, since it gives the fairest outcome from her point of view. However, player $A$ would not consider this outcome to be fair when she compares $p_A = 1$ with $p_B = 0$, since from her point of view, choosing to keep $L$ is something "wrong" to do.

Observe that we have here two *interconnected 1-active-player* games. Denote G-I (respectively G-II) to the game in which only player $A$ (respectively $B$) is active. We say that both games are interconnected because the payoff of the players in each game depend on actions and beliefs held on the other game. Let $b_i^{k,I}$ (respectively $b_i^{k,II}$) be the k-order beliefs held by player $i$. Then, players' payoffs are as follows:

[12] Note that this interpretation of fairness differs from Rabin's (1993) interpretation.

- In G-I: $v_A(a_A; b_A^0(a_A))$; $v_B(a_A)$

- In G-II: $v_A(a_B; b_A^0(a_A); b_A^1(a_B))$; $v_B(a_B; b_B^1(a_A))$

Both G-I and G-II are psycho-social games and G-II is also a GPS's psychological game.

Each active player takes as given her beliefs and compute her best response as in any standard psychological or psycho-social game. In G-I, player $A's$ best response is $a_A^* \in \arg\max v_A(a_A; b_A^0)$ for a given $b_A^0$. In G-II, player $B's$ best response is $a_B^* \in \arg\max v_B(a_B; b_B^1)$ for a given $b_B^1$. In the equilibrium of the combined game (G-I and G-II), $b_A^{*0}(a_A^*) = a_A^*$ and $b_B^{*1}(a_A^*) = b_A^{*0}$.

This example has three psycho-social equilibria:

Type I: Player $A$ believes that the right thing to do is to give player $B$ the large apple ($p_A = 1$), so if she is the active player, she will choose to do so. When player $B$ is the active player, given her beliefs, she will choose to keep the large apple for her ($p_B = 0$). Formally, Type I equilibrium is defined by the following quadruple ($a_A^* = S_A, a_B^* = L_B; b_A^{*0}(a_A^*) = S_A, b_B^{*1}(a_A^*) = b_A^{*0} = S_A$). The equilibrium payoffs are:

- If G-I is played: $(1, 0)$

- If G-II is played: $(-1, 1)$

Type II: Player $A$ believes that the right thing to do is to give player $B$ the small apple ($p_A = 0$), so if she is the active player, she will choose to do so. When player $B$ is the active player, given her beliefs, she will choose to keep the small apple for him ($p' = 1$). Formally, Type II equilibrium is defined by the following quadruple ($a_A^* = L_A, a_B^* = S_B; b_A^{*0}(a_A^*) = L_A, b_B^{*1}(a_A^*) = b_A^{*0} = L_A$).The equilibrium payoffs are:

- If G-I is played: $(1, 0)$

- If G-II is played: $(-1, 1)$

Type III: Player $A$ is "morally indifferent" between keeping the small or the large apple ($p_A = \frac{1}{2}$), so if she is active, she will randomize between her two options. When player $B$ is the active player, she will also randomize ($p_B = \frac{1}{2}$). Formally, Type III equilibrium is defined as follows: $(p_A = \frac{1}{2}, p_B = \frac{1}{2}; b_A^{*0}(a_A^*) = \frac{1}{2}, b_B^{*1}(a_A^*) = b_A^{*0} = \frac{1}{2})$.The expected payoffs in equilibrium are:

- If G-I is played: $(\frac{1}{4}, 0)$
- If G-II is played: $(0, \frac{1}{4})$

Type I equilibrium describes the situation of Sen's example. Moreover, it can be inferred from Sen's example that both players have a different notion of fairness. Player $A$ whose behaviour is based on commitment, believes that player $B$ is being fair when she observes ex-post that $|p_B - p_A| = 0$ or unfair otherwise. On the contrary, player $B$ who sympathizes with $A$, thinks that she is being fair when $|p_B - p_A| \neq 0$ and unfair otherwise.

Now suppose that both players base their choices on sympathy. The payoffs of the game are:

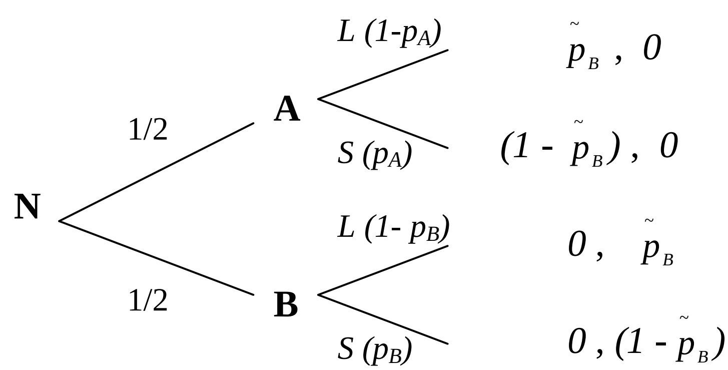


Figure 4.3: Both players base their choices on sympathy

In the situation described by Sen, if player $A's$ choice of the small apple ($p_A = 1$)

was based on "sympathy" instead of on "commitment", player $B$ choosing to keep the large apple ($p_B = 0$) would not be judged by $A$ as being unfair since $|p_B - p_A| \neq 0$.

### 4.4.2 Empowerment and Paternalism

Consider a kid who needs to do a homework for school but he does not know how to do it. He can either ask his mother to help him to understand the problem and then solve the problem on his own (i.e. looking for empowerment), or ask his mother to do the homework for him (i.e. looking for paternalism). His mother can choose either to explain him how to solve the problem without solving it or she could decide to do the homework for him. The set of actions for each player, mother and son, is constituted by $A_i = \{\underline{a}_i, \bar{a}_i\}$ , $\bar{a}_i > \underline{a}_i$,where $\underline{a}_s$ and $\bar{a}_s$ represent the son asking her mother for a paternalist behaviour or for empowerment, respectively, while $\underline{a}_m$ and $\bar{a}_m$ represent the mother choosing a paternalist behaviour or empowering her son, respectively. Let $P_s$ be the set of "self-confidence levels" of the son and $P_m$ be the set of "self-esteem levels" of the mother, where both $P_s$ and $P_m$ are intervals in $\Re$. The payoff of the son is a map $u_s : A_m \times A_s \times P_s$, which depends not only on his own actions (how he manages to have the homework done) but also on his mother's actions and his level of confidence. In addition, there is a map $\pi_s : A_m \times A_s \rightarrow P_s$ which generates the son's self-confidence levels for each configuration of actions of mother and son. Assume that $\pi_s(a_m, a_s)$ is increasing in $a_m$ and $a_s$. In words, the kid will be more self-confident if he does the homework on his own and her mother does not choose to do the homework for him. The payoff of the mother is a map $u_m : A_m \times A_s \times P_m$, which depends not only on her own actions (how she manages to help her son to have the homework done) but also on his son's actions and her level of self-esteem. In addition, there is a map $\pi_m : A_m \times A_s \rightarrow P_m$ which generates the mother's self-esteem levels for each configuration of actions of both, mother and son. Assume that $\pi_m(a_m, a_s)$ is increasing in $a_m$ and $a_s$. In words, the mother will feel more valuable if she helps her son to do the homework on his own and her son does not choose to ask her to do the homework for him. For simplicity, assume that the individual's utility from a paternalistic behaviour $u_i(\underline{a}_i, a_j, p_i)$ is normalized to zero for all values of $a_j$ and $p_i$ and $u_i(\bar{a}_i, a_j, p_i)$ is the gain (or loss) to each individual in deviating from a paternalistic behaviour.

Assume that

*(i)* for each $a_j$, $u_i(\bar{a}_i, a_j, p_i) > u_i(\bar{a}_i, a_j, p_i'), p_i > p_i'$

*(ii)* for each $p_i$, $u_i(\bar{a}_i, a_j, p_i) > u_i(\bar{a}_i, a_j', p_i)$, $a_j > a_j'$

In words this means that $u_i(a_i, a_j, p_i)$ has strictly increasing differences in $a_j$ and $p_i$, i.e. the son's (mother's) marginal return of receiving (providing) empowerment is higher if her mother (her son) chooses to empower him (to be empowered) and if his confidence (her self-esteem) is higher.

Let $\pi_i(\underline{a}_i, a_j) = \underline{p}_i(a_j)$ and $\pi_i(\bar{a}_i, a_j) = \bar{p}_i(a_j)$. Since $\pi_i(a_i, a_j)$ is increasing in $a_i$ it follows that $\bar{p}_i(a_j) > \underline{p}_i(a_j)$.

For each $a_j$ and $p_i$, let $\alpha_i(a_j, p_i)$ be the set of actions that maximize individual $i's$ payoffs.

A psycho-social equilibrium is a pair $(a^*, p^*)$ such that for each individual, $(i)$ given $p_i^*$ and $a_j^*$, $a_i^* \in \alpha_i(a_j^*, p_i^*)$ and $(ii)$ given $a_i^*$ and $a_j^*$, $p_i^* \in \pi_i(a_i^*, a_j^*)$.

What are the psycho-social equilibria of this example?

Under the assumptions made so far, there is a unique solution, $\hat{p}_i(a_j)$, for the equation $u_i(\bar{a}_i, a_j, p_i) = 0$ and due to conditions *(i)* and *(ii)*, $\hat{p}_i(a_j)$ is decreasing in $a_j$. That is, in equilibrium, the level of confidence (self-esteem) that the son (mother) needs in order to choose to be empowered (to empower) is higher the less his mother (her son) chooses to empower (to be empowered) his son (by her mother).

Then, given $a_j$, the best response correspondence of player $i$ is determined as follows:

(i) if $p_i > \hat{p}_i(a_j)$, $\bar{a}_i = \alpha_i(a_j, p_i)$

(ii) if $p_i < \hat{p}_i(a_j)$, $\underline{a}_i = \alpha_i(a_j, p_i)$

(iii) if $p_i = \hat{p}_i(a_j)$, $\{\underline{a}_i, \bar{a}_i\} = \alpha_i(a_j, p_i)$

Since $\pi_i(\underline{a}_i, a_j) = \underline{p}_i(a_j)$ and $\pi_i(\bar{a}_i, a_j) = \bar{p}_i(a_j)$ and since both players are symmetric, then, the psycho-social equilibria of this example are:

(i) if $\hat{p}_i(a_j) > \bar{p}_i(a_j)$, $(a^*, p^*) = (\underline{a}_i, \underline{p}_i(\underline{a}_j))$

(ii) if $\hat{p}_i(a_j) < \underline{p}_i(a_j)$, $(a^*, p^*) = (\bar{a}_i, \bar{p}_i(\bar{a}_j))$

(ii) if $\bar{p}_i(a_j) \geq \hat{p}_i(a_j) \geq \underline{p}_i(a_j), (a^*, p^*) = \Big\{(\underline{a}_i, \underline{p}_i(\underline{a}_j)), (\bar{a}_i, \bar{p}_i(\bar{a}_j))\Big\}$

The most interesting equilibrium of this example is $(\underline{a}_i, \underline{p}_i(\underline{a}_j))$, a *disempowerment* equilibrium. In this situation, the son asks for help and the mother does the homework for him. This behaviour sustains low levels of self-confidence and self-esteem in son and mother.

## 4.5 Nash vs. Psycho-social equilibria

> "...In essence, we constantly ignore Kelley and Thibaut's (1978) long accepted observation that the people in our experiments transform the payoff matrices that we give them, and then they act in ways that maximize their transformed outcomes. Yet we constantly ignore this knowledge, probably so that we don't have to complicate our work too much." (in Murnighan and Roth (2006, pg. 8))

What do people play when they play a normal-form game? Do they actually play the game that the experimenter thinks they play? This is still an open question. The vast experimental evidence showing that players do not behave as it is predicted by Nash is still puzzling. Some scholars argue that people may misunderstand the game. However, experimenters are very careful by making the rules and the payoffs of the game simple and clear in order to minimize any possible misunderstanding. Even in cases in which the simplicity of the game is remarkable (e.g. Ultimatum or Dictator game), the Nash predictions are rarely met. In this chapter we argue that people may be playing "psycho-social" games instead, which provides a theoretical explanation of "out of equilibrium" behaviour in the experimental settings. Indeed, we generalize what Rabin (1993), Charness and Rabin (2002), Charness and Dufwenberg (2006) or Shalev (2000) and others have done. Rabin (1993), in particular, derives a "reciprocity" psychological game from some basic material normal-form games providing a link between both classes of games. However, Rabin's analysis is restrictive to the particular functional form he assumes for the map $\pi_i$. In principle, one could use other sensible functional forms to derive other psycho-social games from the same normal-form game. One single normal-form game can be associated to many psycho-social games, not just a "reciprocity" one.

Consider the following prisoner's dilemma normal-form game with $X > 0$:

| | | Player 2 | |
|---|---|---|---|
| | | C | D |
| Player 1 | C | $4X, 4X$ | $0, 6X$ |
| | D | $6X, 0$ | $X, X$ |

(4.5)

One could in principle derive many sensible (mathematically infinite) psycho-social games from this standard prisoner's dilemma. Here we consider some examples. Let $a_i$ be player $i$ beliefs about what she chooses, $b_j$ player $i$ beliefs of what player $j$ chooses and $c_i$ player $i$ beliefs about what player $j$ believes player $i$ chooses.

**Example 14** ***Fairness-Reciprocity Game****: If $p_i = (b_j, c_i)$ and $v_i(a_i, p_i) = x_i(a_i, b_j) + \rho_i \left[ f_i(a_i, b_j)\tilde{f}_j(b_j, c_i) \right]$ with $\rho_i \geq 0$ being a constant measuring $i'$s sensitivity to reciprocate, $f_i(a_i, b_j)$ being $i'$s kindness towards $j$ and $\tilde{f}_j(b_j, c_i)$ being $j'$s kindness towards $i$, then one can derive a "fairness-reciprocity" game from the game above (see Rabin (1993) for further detail)*

**Example 15** ***Guilt Game:*** *If* $p_i = (c_i)$ *and*

$v_i(a_i, p_i) = x_i(a_i, b_j) - \gamma_i \max\{0, x_j(c_i) - x_j(a_i)\}$, *with* $\gamma_i \geq 0$ *being a constant measuring* $i'$*s guilt aversion,* $x_j(c_i)$ *and* $x_j(a_i)$ *being* $j'$*s material payoffs given* $c_i$ *and* $a_i$ *respectively, then one can derive a "guilt-aversion" game from the game above.*

**Example 16** ***Commitment Game:*** *If* $p_i = (a_i)$ *and*

$v_i(a_i, p_i) = x_i(a_i, b_j) - \alpha_i \max\{0, x_j(a_i) - x_j(a_i)\}$, *with* $\alpha_i \geq 0$ *being a constant measuring* $i'$*s sensitivity to moral commitment and* $x_j(a_i)$ *being* $i'$*s material payoffs given* $a_i$ *respectively, then one can derive a "commitment" game from the game above.*

**Proposition 9** *A cooperative equilibrium (C,C) exists in a any of this three psycho-social games associated with the prisoner's dilemma material game. In particular, (C,C) is a "fairness-reciprocity" equilibrium if* $\rho_i \geq 4X$*, it is a "guilt-aversion" equilibrium if* $\gamma_i \geq \frac{1}{2}$*, and it is a "commitment" equilibrium for any* $\alpha_i \geq \frac{1}{2}$*.*

In general, one can imagine other psycho-social games being derived from any arbitrary material normal-form game. In the remaining of the section, we introduce a general way to associate a normal-form game with a psycho-social game. We show that if we take any normal-form game and we look at the psycho-social equilibria of those psycho-social games associated to it, we shall find that they are generically different from the Nash equilibria of the original normal-form game. Throughout this section we will work just in pure strategies to simplify the notation and to gain intuition, but the analysis shall be extended to mixed strategies in the near future.

### 4.5.1 Embeddedness

Let $\psi := \{A_i, P_i; v_i : A \times P_i \to \Re; \pi_i : A \to P_i\}$ denote any finite n-person psycho-social game as defined in Section 3.3[13]. Let $\psi \in \Psi(A_1, ..., A_n)$ where $\Psi(A_1, ..., A_n)$ is the

[13] To simplify the notation, we shall assume that preferences do not depend on $p_{-i}$ and we shall also rule out the possibility that $p_{-i}$ affects $p_i$.

*set of all finite psycho-social* games with strategy spaces $A_1, ..., A_n$. Now, let $\lambda := \left\{\tilde{A}_i; u_i : \tilde{A} \rightarrow \Re\right\}$ denote any finite n-person standard normal-form game where $\tilde{A}_i$ is a finite nonempty set of pure strategies. Let $\lambda \in \Lambda(\tilde{A}_1, ..., \tilde{A}_n)$ where $\Lambda(\tilde{A}_1, ..., \tilde{A}_n)$ is the *set of all finite standard normal form* games with strategy spaces $\tilde{A}_1, ..., \tilde{A}_n$.

**Definition 4** ***Embededdness***: *Take any $\lambda$ from the set $\Lambda$ and any $\psi$ from the set $\Psi$. We say that $\lambda \in \Lambda$ is consistently embedded into $\psi \in \Psi$ (with the following notation: $\lambda \overrightarrow{e}\ \psi$) if, for all $i \in N$, (1) $A_i \equiv \tilde{A}_i$ and (2) $v_i(a, p_i) = u_i(a)$ for any $p_i = \pi_i(a)$ and all $a \in A$.*[14]

There are two other - rather extreme - ways to define a normal form game embedded into a psycho-social game, although we decide to work with the "consistently embedded" definition for two main reasons. First, we could have said that $\lambda$ is "weakly embedded" into $\psi$ if, for all $i \in N$, condition (1) and (2) hold for any $p_i \in P_i$, which does not need to be a consistent $p_i$. We rule-out this restriction since it does not have much intuition behind. Further, we could have said that $\lambda$ is "strongly embedded" into $\psi$ if for all $i \in N$, condition (1) and (2) hold for some $p_i = p_i^* \in P_i$. If we consider this definition, we are imposing the overly-strong restriction that the $\psi$ has at least one psycho-social equilibrium in pure strategies, and then we rule out games without pure strategy psycho-social equilibrium.

Now we have defined a general way to associate a normal form game to a psycho-social game, we shall compare the set of equilibria of both classes of games.

**Proposition 10** *Let $NE$ be the set of Nash equilibrium strategy profiles of $\lambda \in \Lambda$ and $PSE$ be the set of psycho-social equilibrium strategy profiles of $\psi \in \Psi$. If $\lambda \overrightarrow{e}\ \psi$ then generically, under standard regularity assumptions, both sets of equilibria in pure strategies are distinct from each other.*

[14] Note that there is no technical reason for which we have to imposse "payoff equivalece" between the normal form game and the projected psycho-social game. It would be sufficient to state that $a^* = \tilde{a}^*$, i.e. both are strategically equivalent. However, we prefer to work with "payoff equivalence" because we think that it simplifies the notation and it is innocuous.

Both, the definition of embededdness and proposition 10 consider equilibria in pure strategies and interior solutions. If we wanted to allow for mixed strategies, then the first order conditions used for the proof would not be valid any more.

### 4.5.2 Restricted Psycho-Social Games

In a psycho-social game, there is a feedback effect arising from the interaction in the game which is not internalized by the players when they compute their best reply. This feedback is modelled by adding an extra endogenous argument on players' utility function, thus broadening their preference domain. The intuition is that besides the material payoffs of the game there is an extra argument (emotions, beliefs, aspirations, values) which players care about, and that is developed within the same interaction. However players do not perfectly reflect on that.

There is another way of modelling the same notion of a feedback effect not being internalized by the players. Suppose that, instead of broadening the preference domain, we restrict it with an endogenous parameter. If the players compute their best response using the restricted utility, they will impose an externality on themselves without internalizing it. This second approach is closer to the bounded rationality idea of Simon (1959) than our first approach. Simon claims that economic agents strive to maximize their utility and instead they employ heuristics to make decisions. They do this because their inability to process and compute the expected utility of every alternative action "even in an extremely simple situation" (Simon, 1959, pg. 261)

Note that the two approaches differ in the degree of reflection imposed on the players. In the augmented approach, a non-fully reflective player does not internalize the fact that she can affect, for example, her emotional reactions, while a fully reflective player does. In that sense, a fully reflective player in the augmented approach is too reflective, i.e. he is cognitively and emotionally intelligent. However, the degree of reflection required by the restricted approach is lower: a fully reflective player is only required to be cognitively intelligent.

In what follows, we study some of the implications for game theory of assuming the second approach. We show, for example, that a strict Nash equilibrium is robust to any kind of imperfect reflection. Even assuming the highest level of irreflection, if a normal-form game has a unique strict Nash equilibrium, then this equilibrium remains regardless how irreflective the players are. As stated by Aumann (1997, pg. 4), "one of the simplest, yet most fundamental ideas in bounded rationality (indeed, in game theory as a whole) is that no rationality at all is required to arrive at a Nash equilibrium." Proposition 11 shows that this statement is true in our setting, although only in cases in which there is a unique strict pure strategy Nash equilibrium. Proposition 12 shows that some degree of rationality is needed in normal-form games with mixed strategy equilibrium.

### Model

Let $N = \{1, ..., n\}$ be the set of players, and for each $i \in N$, let $A_i$ be the nonempty, finite set of actions available to player $i$. For any set $X$ (where the topology of $X$ is understood), $\Delta(X)$ denotes the set of (Borel) probability measures on $X$. Thus, $S_i := \Delta(A_i)$ is the set of mixed strategies of player $i$. Let $S := \Pi_{i \in N} S_i$ and $S_{-i} := \Pi_{j \neq i \in N} S_j, i \in N$. Each strategy profile $s \in S$ induces a probability distribution over the outcome set (or set of pure strategy profiles) $A := \Pi_{i \in N} A_i$.

Player $i's$ utility function is $x_i : A \rightarrow \Re$, and the expected value of $x_i$ is

$$v_i(s_i, s_{-i}) := \sum_{a_{-i} \in A_{-i}} \sum_{a_i \in A_i} s_{-i}(a_{-i}) s_i(a_i) x_i(a_i, a_{-i}) \tag{4.6}$$

Given $s$, $A'_i \subset A_i$, $A'_{-i} \subset A_{-i}$, let

$$v_i(s_i, s_{-i}|\, A'_i, A'_{-i}) := \sum_{a'_{-i} \in A'_{-i}} \sum_{a'_i \in A'_i} s_{-i}(a_{-i}) s_i(a_i) x_i(a_i, a_{-i}) \tag{4.7}$$

**Definition 5** *$v_i(s_i, s_{-i}|\, A'_i, A'_{-i})$ is a restriction of $v_i(s_i, s_{-i})$ for $A'_i$ and $A'_{-i}$. A restriction is consistent across $i$ if and only if, for all $i \in N$, $A'_{-i} = \Pi_{i \in N} A'_i$.*

**Definition 6** *A feedback $\pi_i(s_i, s_{-i})$ is a restriction $v_i(s_i, s_{-i}|\, A'_i, A'_{-i})$ for some $A'_i \subset A_i$, $A'_{-i} \subset A_{-i}$. A feedback is non-trivial if either (i) $A'_i \neq \varnothing$ or $A'_{-i} \neq \varnothing$ or both, or (ii) $A'_i \times A'_{-i} \neq A$, otherwise the feedback is trivial. A profile of feedbacks $\pi := (\pi_i : i \in N)$ is consistent if the underlying profile of restrictions is consistent.*

**Definition 7** *A game $G := \{A_1, ..., A_n; v_1, ..., v_n\}$ is finitely parametrizable if and only if there exists $u_i : S \times \Re^K \to \Re$ and $\delta_i^k : S \to \Re$ such that $v_i(s_i, s_{-i}) = u_i(s_i, s_{-i}; \delta_i^1(s_i, s_{-i}), ..., \delta_i^K(s_i, s_{-i}))$, $\forall i \in N$.*

**Definition 8** *A restricted normal form game $G' := \{A_1, ..., A_n; v_1, ..., v_n\}$ is a finitely parametrizable game with $\delta_i(s_i, s_{-i}) = \pi_i(s_i, s_{-i})$, where $\pi_i(s_i, s_{-i})$ is a feedback.*

**Definition 9** *An equilibrium of $G$ is a pair $(\hat{s}, \hat{\delta}) \in S \times \Re^K$ such that:*
*(i) $u_i(\hat{s}_i, \hat{s}_{-i}; \hat{\delta}_i^1, ..., \hat{\delta}_i^K) \geq u_i(s_i, \hat{s}_{-i}; \hat{\delta}_i^1, ..., \hat{\delta}_i^K), \forall i \in N, \forall s_i \in S_i$, $\forall k \in K$ and*
*(ii)$\hat{\delta}_i^k = \delta_i^k(\hat{s}_i, \hat{s}_{-i}), \forall i \in N, \forall k \in K$.*

**Definition 10** *Any two games $\Gamma := \{A_1, ..., A_n; v_1, ..., v_n\}$ and $\hat{\Gamma} := \{A_1, ..., A_n; \hat{v}_1, ..., \hat{v}_n\}$ are best-response equivalent if and only if:*
$\arg\max_{a_{-i} \in A_{-i}} \sum_{a_{-i} \in A_{-i}} s_{-i}(a_{-i}) x_i(a_i, a_{-i}) = \arg\max_{a_{-i} \in A_{-i}} \sum_{a_{-i} \in A_{-i}} s_{-i}(a_{-i}) \hat{x}_i(a_i, a_{-i}), \forall i \in N, \forall s_{-i} \in S_{-i}$

**General Results**

Can any Nash Equilibrium of a finitely parametrizable game be preserved as an equilibrium of any restricted normal form game? The answer is Yes, at least when the Nash Equilibrium is strict in pure strategies. This result is stated in the following proposition.

**Proposition 11** *Take any finitely parametrizable normal-form game* $\Gamma$ *and let* $a^* = (a_i^*, a_{-i}^*)$ *be a pure strategy Nash Equilibrium profile of* $\Gamma$*. There exists a "best-response" equivalent normal form game* $\hat{\Gamma}$ *whose set of equilibria contains* $a^* = (a_i^*, a_{-i}^*)$ *under any restriction.*

This result may not hold if the normal form game has multiple equilibria in pure strategies. As an example, take the following normal form game:

| | L | R |
|---|---|---|
| U | $\underline{4}$,$\underline{4}$ | -2,3 |
| D | 3,-2 | -$\underline{1}$,-$\underline{1}$ |

(4.8)

with $(U, L)$ and $(D, R)$ being the two Nash equilibrium in pure strategies of the game. It easy to show that there is no strategically equivalent game such that both equilibria can be preserved for any restriction

Proposition 11 holds for normal form games because of their bilinearity. However, this is not necessarily true for any simultaneous move games.

The result stated in Proposition 11 assumes Nash Equilibrium in pure strategies. In Proposition 12, we show that we cannot claim an analogous result if the Nash Equilibrium is in mixed strategies. Hence, an equilibrium in mixed strategies of a finitely parametrizable normal form game will not necessarily be preserved as an equilibrium of any (non-trivial) restricted normal form game.

**Proposition 12** *Take any finitely parametrizable non-trivial normal form game* $\Gamma$ *and a non-trivial non-degenerate mixed strategy Nash Equilibrium* $\Gamma$, $s^* = (s_i^*, s_{-i}^*)$. *There is no strategically equivalent transformation of* $\Gamma$ *that preserves* $s^*$ *as a mixed strategy equilibrium of the transformed restricted game for any non-trivial restriction.*

Of course a non-degenerate mixed strategy equilibrium is preserved for some restrictions, but in Proposition 12 we claim that this is not true for all.

Given the generality of our model, we ask ourselves if any outcome of any game could be rationalized as a psychosocial equilibrium for some appropriate restriction. To answer this question, it is enough to show that there exists at least one possible outcome of a game that is not a psyschosocial equilibrium for any non-trivial restriction. In the next proposition, we show that such an example exists, and therefore, the model is testable.

**Proposition 13** *Not every outcome of any game can be rationalized as an equilibrium of any restricted game for any non-trivial consistent restriction.*

## 4.6 Imperfect vs Perfect Reflection

This section explores the relevance of the assumption of imperfect reflection in some existing games with endogenous preferences. We find different results depending on the game we explore. For instance, in Rabin (1993), no matter how reflective the players are, the predictions of the games do not change. The same is true in Shalev (2000). However, in other games (e.g. with guilt or commitment) the assumption of imperfect reflection changes the predictions of the game.

### 4.6.1 Imperfect Reflection in Social Preferences

How does the "fixed $p_i$" assumption works in the endogenous social preferences models? The existing models on endogenous social preferences assume imperfect reflection in players. However, it turns out that the particular specification of Rabin (1993) makes this assumption innocuous - i.e. the set of equilibria are the same under any of the two assumptions. It is not the case, however, with "guilt-aversion" and "commitment" models.

Consider the following three motivations for social preferences: reciprocity, guilt and commitment. If $(a^*, p^*)$ is a psycho-social equilibrium profile of a "fairness with reciprocity" game, "guilt" game and "commitment" game, then given $p^*$, $v_i(a_i^*;p_i^*) \geq v_i(a_i';p_i^*)$ and $p_i^* = \pi_i(a_i^*)$ for each player $i$. In particular:

a) in a "fairness with reciprocity" game (Rabin, 1993), player $i$ does not have individual incentives to deviate from playing $a_i$ iff $v_i(a_i;b_j, c_i) \geq v_i(a_i';b_j, c_i)$. When an irreflective reciprocal player $i$ computes the gains from deviating, she re-computes $f_i(a_i', b_j)$ but she leaves $\tilde{f}_j(b_j, c_i)$ unchanged, i.e. she does not fully internalize the consequence of her deviation on her set of beliefs because $c_i$ remains fixed. However, if the reciprocal player had perfect reflection, she would have incentives to deviate from playing $a_i$ iff $v_i(a_i;b_j, c_i) \geq v_i(a_i';b_j, c_i')$ and $a_i' = c_i'$.[15] Then, given the new set of beliefs $p_i = (b_j, c_i')$, a reflective and reciprocal player $i$ will re-compute $f_i(a_i', b_j)$ and $\tilde{f}_j(b_j, c_i')$ and evaluate the gains from deviating. It turns out that given the particular specification of Rabin (1993) $\tilde{f}_j(b_j, c_i') = \tilde{f}_j(b_j, c_i)$, so the set of equilibria assuming either sophisticated or myopic players is the same. Intuitively, it seems more reasonable to think that $\tilde{f}_j(b_j, c_i') \neq \tilde{f}_j(b_j, c_i)$. Going back to the prisoner's dilemma example, one would expect (as for one I do) that if player $j$ is a reciprocal player, she would be less kind with player $i$ if she believes player $i$ will defect $(c_i = D)$ when she cooperates $(b_j = C)$. However, Rabin's specification implies that player $i$ thinks that $j$ is equally generous to him when she cooperates, whatever his beliefs are about what player $i$ will do. This contra-intuitive result seems to highlight a failure in the specification of Rabin's model. Intuitively, one would expect $\tilde{f}_j(C, D)$ to be higher than $\tilde{f}_j(C, C)$, but because it is not the case, Ra-

[15] Note that first order beliefs $b_2$ always are kept fixed when evaluating a deviation, as in any standard game.

bin's behavioural model does not have different normative implications from a standard model. The descriptive and normative implications of Rabin's model are robust to the degree of reflection of the players (in the sense that they do not update their beliefs when they play the game).

b) In "guilt" and "commitment" games, it is clear that the assumption of myopia may change the set of equilibria (see Example 15, pg. 93, for a "guilt" game). Consider the pure "commitment" game in figure 4.1 (pg. 84). In that example, if the player were reflective enough, she would solve the following problem

$$\underset{p_i}{Max}(1-p_i)(1-\tilde{p}_i)+p_i\tilde{p}_i \quad st.p_i=\tilde{p}_i, p_i\in[0,1] \quad \Leftrightarrow \tag{4.9}$$

$$\underset{p_i}{Max}(1-p_i)^2+p_i^2 \;\text{ s.t. }\; p_i\in[0,1] \tag{4.10}$$

which has two corner solutions or equilibria: $p_i=\tilde{p}_i=0$ and $p_i=\tilde{p}_i=1$ (both with payoff of 1). Note that there is no mixed strategy equilibrium in the perfect reflection case. However, the mixed strategy psycho-social equilibrium exists (with payoff of $\frac{1}{2}$) even if we assume irreflective players.

### 4.6.2 Imperfect Reflection in Loss-Aversion Games

Besides defining a "myopic" loss-aversion equilibrium, Shalev (2000) also defines a different concept of equilibrium called *non-myopic loss aversion equilibrium*. Formally, a strategy profile $a$ is defined to be a *non-myopic loss aversion equilibrium* if for all $i\in N$ and for all $a_i'\in A_i$, $p_i(a)\geq p_i(a_{-i},a_i')$.

A *non-myopic* loss aversion equilibrium is one that assumes perfect reflection. Players are assumed to take into account the change on her reference point that is consistent with her deviation. Proposition 14 states that myopic loss-aversion games are indistinguishable from non-myopic loss aversion games.

**Proposition 14** *Under approach 1, every (pure strategy) Nash equilibrium of the original game is both loss aversion psycho-social equilibrium and non-myopic loss aversion equilibrium.*

## 4.7 Conclusion

We have introduced a general class of simultaneous move games in which the payoff of each player depends not only on her strategy profile, but also on her preference parameters. The preference parameters are, in turn, endogenously determined in equilibrium. When these games are played by players without perfect reflection, they are labelled "psycho-social games".

We linked our framework to the existent literature on social preferences and developed two applications of psycho-social games introducing the concept of individual commitment and empowerment into game theory.

We proved equilibrium existence under weak assumptions, extending the existence proof of Chapter 2. We associated psycho-social games to standard normal-form games in two different ways. We showed that, typically, the set of Nash equilibria and the set of psycho-social equilibria of an associated psycho-social game are distinct from each other. We have also shown that a strict Nash equilibrium is robust to all degrees of imperfect reflection. This result is consistent with Aumann's (1997) claim that "no rationality at all is required to arrive at a Nash equilibrium" (p. 4). This is true when the game has a unique strict pure strategy Nash equilibrium. If the normal-form game has weak or mixed strategy equilibria, then we show that some degree of "rationality" in Elster's (1983) broad sense is needed to sustain a weak Nash equilibrium

# Chapter 5

# Poverty Persistence and Aspirations Failure

## 5.1 Introduction

In contrast to transient poverty, persistent poverty[1] is not just a snapshot of those who are poor now, but a condition that implies an understanding of a multidimensional process which makes people poor and keeps them poor[2].

An influential literature, advocated mostly by economists studying the origins of "poverty traps," argues that persistent poverty is due to constraints that are *external* to the individual.[3] Examples of such constraints are credit or insurance market imperfections (Loury, 1981; Galor and Zeira, 1993; Banerjee and Newman, 1991, 1993) coordination problems (Da Rin and Hellmann, 2002; Kremer, 1993), institutional or governmental failures (Bardhan, 1997), malnutrition (Dasgupta and Ray, 1986), neigh-

[1] Persistent poverty is defined by the incapability to fulfil basic needs during a period greater than 5 years,

[2] More than 300 million people world-wide have lived in persistent poverty in the late 1990s. The Chronic Poverty Report 2004-2005 estimates that 40% of the poverty in Sub-Saharan Africa is persistent. For evidence of chronic or persistent poverty see Jalan and Ravallion (1998); Fouarge and Layte (2005); Biewen (2006); Duncan et al. (1993), among others.

[3] See for example Azariadis (2004); Azariadis and Stachurski (2005) for a literature review on Poverty Traps

borhood effects (Durlauf, 2006) or fertility decisions (Nelson, 1956).

A different approach argues that a poverty trap is perpetuated by the interaction between extrinsic circumstances (initial disadvantage e.g. poverty or social exclusion) and *intrinsic factors* such as aspirations and beliefs. The sole fact of being poor may affect preferences by constraining aspirations, self-confidence or hope which may in turn limit poor people's ability to participate and alter their own condition.[4] "Long-run poverty is fundamentally self-perpetuating [and] the entrapment goes hand in hand with [...] lack of hope," argues Mookherjee and Ray (2003, pg. 5).

As pointed out by the anthropologist Appadurai (2004, pg. 59), poor people may lack the capacity to aspire to "contest and alter the conditions of their own poverty." For Appadurai, the "capacity to aspire" is a navigational capacity which includes not only the ability to set goals or aspirations but also the knowledge of how to achieve them. The higher the initial advantages, the more chances an individual has to set higher aspirations and to see the pathways which lead to their fulfillment. In contrast, the precarious life of poverty restricts the poor's aspiration levels to those of necessity, reinforcing and perpetuating the cycle of poverty and aspiration failure. A possible way out of this trap, in Appadurai's view, is to expand the poor's aspiration horizon by means of role models or to create programs which provide the poor with an arena in which to develop capabilities and have voice. Thus the capacity to aspire is a key ingredient in any notion of empowerment[5].

[4] Social psychologists have documented the lack of hope and low self-esteem as a typical endogenous characteristic in the personality of the poor population. Moreira (2006) studies the poor people in the North-eastern Brazil and points out that "as the poor lose their values, they no longer believe in themselves. They go through a process of Nihilism [denial of hope]". He documents evidence that shows that the greatest part of the poor population has these nihilistic characteristics and they submit themselves to the destiny that is given by God. The work of the renowned sociologist W. J. Wilson provides a clear evidence of the link "social exclusion-lack of aspirations-poverty" that was observed in U.S. urban ghettos since 1970. Wilson (1987) makes the case that the increasing "social isolation" of the poor, especially the black poor, has greatly contributed to their poverty. "Out of sight, out of mind" allowed most of the non-poor to either deny or forget the conditions in the ghetto. Moreover, Wilson argues that the causality takes the other direction as well: poverty also implies exclusion. He claims that concentration of poverty results in the isolation of the poor from the middle class and its corresponding role models, resources, and job networks. More generally, he argues that being poor in a mixed-income neighbourhood is less damaging than being poor in a high poverty neighbourhood.

[5] Atkinson (1998) defines social exclusion as a relative concept (people are excluded from a particular

Much of classical Economic Theory makes a sharp distinction between preference parameters (aspirations and beliefs) and their external circumstances (initial status or endowments) and assumes that these are both exogenous to the individual. If there is a systematic link between initial disadvantage and aspirations and beliefs, one could directly *assume* that preference parameters and external circumstances are jointly determined so that initial disadvantage and aspirations failure are inextricably linked. Such an approach, however, requires the assumption that individuals in a state of persistent disadvantage are intrinsically different from all other individuals: persistent poverty is due to *innate* personal traits (personality characteristics, preferences etc.) and therefore policy intervention can make little or no difference. From the perspective of analysis, such an approach implies that the economist can describe but cannot explain the ways in which extrinsic disadvantage and internal constraints interact.

This chapter introduces a simple formal model that *derives* the link between initial disadvantage with aspirations, beliefs and choices[6] in a tractable way while yielding testable implications that can be checked against data collected under controlled conditions.

We consider a decision-maker who is endowed with an initial status and who has to choose between two actions: one that perpetuates her initial status and one that has the potential of improving her initial status.[7] The decision-maker's initial status plus her aspiration level defines the target status, i.e. how far in her status she wants to go. The final (realized) status is determined through a production function in the following way. If the decision-maker chooses the action that perpetuates her status-quo, her final status will be equal to her initial status (with probability 1). If, on the contrary, the decision-

society) that involves agency (people may exclude themselves) and it implies future hopes and expectations. People are excluded not just because they are currently without a job or an income but because they have little prospects for the future.

[6] Our model can be viewed as a contribution to fill in a gap in the literature pointed out by Duflo (2006, pg. 10): "... what is needed is a theory of how poverty influences decision-making, not only by affecting the constraints, but by changing the decision-making process itself".

[7] The status of an individual can be broadly interpreted as her income level, education level, wealth, standard of living, location, type of occupation, etc. Moreover, it does not need to be material. The individual might want to change her political power, cultural status or even her recognition.

maker chooses an action that has the potential of improving her status, she will reach her target with probability $p$ and stays with her initial status with probability $1-p$ where $p$ is a combination of external signals (for example a role model) and her subjective beliefs (self-confidence). The decision maker gets a benefit from the final status achieved and there is a cost of action which depends on her initial status. Choosing the action that perpetuates the status-quo yields no benefit and is costless. Any attempt to change the status-quo is costly but (a) the cost of doing so is lower the higher is the initial status (so that the initial status is a measure of the level of disadvantage of the decision-maker) and (b) the target status is achieved with probability $p$ and the benefit function is increasing in the target status (and hence, the aspiration level).

We first assume that initial status, aspirations and probabilities are given to the individual. We show that initial advantage and self-confidence are substitutes for aspirations. That is, the more disadvantaged the decision-maker is (respectively the lower self-confidence), the higher aspirations she needs in order to find it convenient to undertake an action that has the potential to change her status-quo. The main implication of this result is that initial disadvantage is associated with aspirations failure, a combination of low aspirations and low achievement.

Next, we allow for endogenous aspirations. We study the impact of a role model as a way out of a deprivation trap. A role model provides direct information about $p$ through a successful experience of a "similar" other. As in Gilboa and Schmeidler (2001),[8] we endow the decision-maker with a "similarity function"[9] which provides a quantification of her subjective assessment of how similar the initial status of the role model is relative to her own initial status. The more similar the role model is to the decision-maker, the stronger will be the positive impact of the role model on her beliefs,

[8] Gilboa and Schmeidler study Case-Based Decision Theory, people make decisions by analogies to past cases, choosing acts that performed well in the past in "similar" situations, and avoiding acts that performed poorly.

[9] We argue that the simillarity functions used in our paper are related to certain sample estimators used in non-parametric econometrics.

and the more likely will it be to break an aspiration failure.[10] We then define the decision-maker's "cognitive neighborhood" as the set of people from where she draws a role model (e.g. peers, parents, etc.). The limits of the neighborhood are chosen optimally by the decision-maker, and it is assumed there is a "cognitive cost" associated with it. That is, the bigger the neighborhood, the more costly is to process the information. It is shown that an initially disadvantaged decision-maker will find it optimal to restrict the size of her "cognitive neighborhood" to those who are similar to her. We use this result to model how overlapping cognitive neighbourhoods lead to an information externality that determines the internal constraints an individual imposes on herself and construct an example to show how aspirations failure arises in a polarized society.

Finally, we study aspirations formation with adaptive preferences. We introduce a feedback effect from actions and initial status to aspirations, and consider two different decision processes. One, in which the decision-maker is sophisticated enough to internalize the feedback effect when choosing an optimal action. We label such decision-problem a "standard decision problem" (SDP). In this case, we show that initial disadvantage is associated with aspirations failure and could result in a poverty trap. Alternatively, with some degree of "myopia", the decision-maker may not internalize the effect of actions into his aspirations, and make his optimal decision as if his aspirations where exogenously given, although both actions and aspirations are required to be mutually consistent. We call such a problem, a Behavioral Decision Problem (BDP). In a BDP, there are two types of equilibria. For individuals with low initial status, the equilibrium is to remain in status-quo with low aspirations ("aspiration failure equilibrium"). For those with high initial status, the equilibrium is to change and sustain high aspirations. For individuals in the middle, with intermediate values of initial status, there are multiple welfare ranked equilibria. We show that in this case, both aspirations and choices, via equilibrium selection, can be determined as a (stochastic) function of the individual's

[10] As Ray (2006, pg. 7) argues: "looking at the experience of individuals similar to me is like running an experiment with better controls, and therefore has better content in informing my decisions and my aspirations."

extrinsic circumstances. We show how, with overlapping cognitive windows, the information externality generated by a role model determines which equilibrium is chosen in a behavioral decision problem.

Overall, our analysis suggests a view of poverty traps that relies in an essential way on the interaction between extrinsic circumstances and intrinsic factors. It is important to remark that this approach does not only highlights the effect of initial disadvantage on people's aspirations and self-confidence, but more importantly, it points out that initial disadvantages may have a permanent effect on people's life through its effect on psychological traits. Our analysis implies that empowerment (the combination of opportunity and the "capacity to aspire") is essential to the process by which individuals exit poverty traps.

The rest of the chapter is structured as follows. Section 5.2 reviews the related literature. Section 5.3 develops the model. Section 5.4 introduces the role model effect and a cognitive neighborhood. Section 5.5 endogenizes aspirations. Section 5.6 discusses some policy implications and Section 7 concludes.

## 5.2 Relation to the Literature

Based on Appadurai's (2004) work, the closest work to ours is Ray (2006), who argues that poverty and failure of aspirations may be reciprocally linked within a self-sustaining trap. Ray (2006) also combines a discussion about the way aspirations may be formed and the way they may affect behavior. By providing a story in which individuals choose a level of effort to minimize their aspirations gap, Ray intuitively suggests that individual investment efforts should be minimal for both high and low aspiration gaps. However, no formal framework is provided in that paper.

Heifetz and Minelli (2006) is another close related paper. They study a model of aspiration traps where an individual in period $t = 0$ makes a choice which will affect her attitude for the rest of her life. Heifetz and Minelli's (2006) paper does not study the effect of role models, initial disadvantage or the way in which aspirations are formed as we do here.

This chapter is also related to the literature of relative deprivation because the concept of poverty we use is a relative one.[11] It is also engaged with the literature of empowerment, since it provides an analytical framework to support the theoretical ideas of empowerment.[12]

We borrow from the literature of neighborhood effects (Durlauf, 2006) the idea that persistence in economic status is generated by group-level influences on individuals, although we incorporate the aspiration channel.

The idea that people have aspirations that want to satisfy is not new in the Economics literature. For instance, the literature of reinforcement learning in game theory shows how agents can adapt by comparing payoffs achieved from actions chosen in the past within an aspiration level. (see Bendor et al., 2001; Karandikar et al., 1998).

Finally, this chapter is related to some particular models that take identity issues into consideration to study poverty and social exclusion, such us Akerlof and Kranton (2000), Hoff and Pandey (2004) or Hoff and Sen (2006)

## 5.3 The model

### 5.3.1 Set-up

In the model, we consider a decision-maker who has to choose between two actions: one that perpetuates her initial status and one that has the potential of improving her initial status.

[11] See Stark and Taylor (1991) for a general introduction to relative deprivation, Deaton (2001) for an application to health, Stark (2006) for an application to growth. For empirical papers on relative deprivation, see for example Duclos and Grégoire (2002).

[12] See for instance Stern (2004), World Bank (2002), or Alsop and Heinsohn (2005)

The initial status of the decision-maker is denoted by $\theta_0$ (with the set of all possible initial social status denoted by $\Theta$). The decision-maker aspires to achieve a new social status and we denote the target status by $\widehat{\theta} = \Delta + \theta_0$, $\Delta \geq 0$, $\Delta \in [0, K]$ where $K$ is some large but finite number. We interpret $\Delta$ as a measure of how high decision-maker aspires. Low aspirations correspond to values of $\Delta$ close to zero while high aspirations correspond to values of $\Delta$ close to $K$. In what follows, we will refer to $\Delta$ as the aspiration level of the decision-maker.

The decision-maker has to choose an action $a \in \{\underline{a}, \overline{a}\}$ where $\underline{a}$ denotes the action that perpetuates her initial status while $\overline{a}$ denotes an action that has the potential of changing her initial status in a manner described below.

Let the final or achieved status be denoted by $\theta_1$. The production function for $\theta_1$ is specified as follows:

$$
\begin{aligned}
a &= \underline{a} \Rightarrow \theta_1 = \theta_0 \text{ with probability } 1, \\
a &= \overline{a} \Rightarrow \begin{cases} \theta_1 = \theta_0 \text{ with probability } 1-p \\ \theta_1 = \widehat{\theta} \text{ with probability } p \end{cases}
\end{aligned} \tag{5.1}
$$

where $p$ is the probability that the decision-maker attaches to different outcomes. In general, $p$ will be a combination of objective data (the decision-makers own experience, the signals she observes based on the experience of others (role models)) and prior subjective beliefs (the degree of confidence or optimism of the decision-maker). In what follows, both interpretations of $p$ will be used.

The payoffs to the decision maker is given by $b(\theta_1) - c(a, \theta_0)$ where $b(\theta_1)$ denotes the benefit the decision-maker obtains at the achieved social state and $c(a, \theta_0)$ is the cost of action which depends on the initial status of the individual. We rewrite these payoffs as $b(\Delta + \theta_0) - c(a, \theta_0)$ and make the following assumption:

**Assumption 1**

*(i)* $b(\Delta + \theta_0)$ is continuous and increasing in $\Delta$ with $b(\theta_0) = 0$: *perpetuating the*

*status-quo yields no benefit and the benefit function is increasing in aspirations;*

*(ii)* $c(\underline{a}, \theta_0) = 0$ for all $\theta_0$: *perpetuating the status-quo is costless.;*

*(iii)* $c(\overline{a}, \theta_0) = \overline{c}(\theta_0) > 0$ and continuous and decreasing in $\theta_0$: *any attempt to change the status-quo is costly but the cost of doing so is lower the higher is the initial status. In this way, the initial status is a measure of the level of disadvantage of the decision-maker.*

### 5.3.2 Optimal Actions with Exogenously Generated Aspirations

In this subsection, we take the decision-maker's aspirations, subjective beliefs and initial status as given. The individual will choose $a = \overline{a}$ if and only if

$$pb(\Delta + \theta_0) - \overline{c}(\theta_0) \geq 0 \tag{5.2}$$

Suppose the initial $\Delta$ was picked at random from $[0, K]$ (where $K$ is some large but finite number) according to some continuous pdf $g(p)$ (with associated cdf $G(p)$). Next, given the $p$,$\Delta, \theta_0$, the individual chooses $a \in \{\underline{a}, \overline{a}\}$.

The following result characterizes the optimal choice of the decision-maker:

**Proposition 15** ***(Optimal action as a function of*** **p** ***and*** $\theta_0$***)*** *Suppose payoffs satisfy assumption 1. Then, there exists a threshold level of aspirations* $\hat{\Delta}(p, \theta_0) > 0$ *such that that the decision-maker attempts to change her initial social status iff* $\Delta \geq \hat{\Delta}(p, \theta_0)$*; moreover the threshold* $\hat{\Delta}(p, \theta_0)$ *is decreasing in both* $p$ *and* $\theta_0$*. The probability that the decision-maker will choose* $\overline{a}$ *is* $1 - G(\hat{\Delta}(p, \theta_0))$ *which is increasing in both* $p, \theta_0$ *as well.*

Proposition 15 shows that initial advantage and self-confidence are substitutes for aspirations. That is, the more disadvantaged the decision-maker is (respectively the lower self-confidence), the higher aspirations she needs in order to find it convenient to undertake an action that has the potential to change her status-quo.

Note that by assumption 1, both $\theta_0$ (the initial status of the decision-maker) and $p$ (the probabilities that the decision-maker attaches to different outcomes) are key characteristics of the decision-maker. The main implication of Proposition 1 is that initial status disadvantage and low level of self-confidence is positively associated with low aspirations and as a consequence, low achievement. Specifically, (i) holding initial status fixed, a person with a low $p$ (low level of confidence) requires a higher level of aspirations than a person with a high $p$ (high degree of confidence) in order to undertake an action that changes her status quo, (ii) holding beliefs fixed, a person with a low $\theta_0$ (initial disadvantage) require a higher level of aspirations than a person with a high $\theta_0$ (initial advantage) in order to undertake an action that changes his status quo.

Let $p_0$ denote the prior of the decision-maker which, we assume, the decision-maker takes as given. Suppose $p_0$ is picked at random from $[0, 1]$ according to some continuous pdf $f(p)$ (with associated cdf $F(p)$). Next, given the $p_0$,$\Delta, \theta_0$, the individual chooses $a \in \{\underline{a}, \overline{a}\}$. The following proposition characterizes the optimal choice of the decision-maker:

**Proposition 16** ***(Optimal action as a function of*** $\Delta$ ***and*** $\theta_0$***)*** *Suppose payoffs satisfy assumption 1. If* $p_0$ *is picked from* $[0, 1]$ *according to some continuous pdf* $f(p)$ *(with associated cdf* $F(p)$*), the probability that the decision-maker will choose* $\overline{a}$ *is* $1 - F(\hat{p}(\Delta, \theta_0))$ *which is increasing in both* $\Delta$ *and* $\theta_0$*.*

Note that there is no guarantee that $\hat{p}(\Delta, \theta_0) \leq 1$. If $\hat{p}(\Delta, \theta_0) > 1$, the decision-maker will never choose $\overline{a}$.

The main implication of Proposition 16 is that initial status disadvantage and low aspiration levels is positively associated with low self-confidence and as a consequence, low achievement. Moreover, initial advantage and aspirations are substitutes for confidence.

Consider a simplification of the model by assuming that there are only two levels of aspirations $\{0, \Delta' > 0\}$.

Suppose the decision-maker chooses a $\Delta$ from $\{0, \Delta' > 0\}$ taking into account the fact that by changing $\Delta$ she affects her own choice. The following result characterizes the conditions under which aspirations failure emerges endogenously in such a situation:

**Proposition 17** ***(aspirations failure)*** *If* $p_0 < \hat{p}(\Delta', \theta_0)$, *the decision-maker will optimally choose* $\Delta = 0$.

Proposition 17 is telling us that even if the person is rational enough to anticipate the effect of ther aspirations on her choices, for a sufficiently low level of initial background or self-confidence, it is optimal for the person to choose to have no aspirations to change her status-quo.

## 5.4 Role models, Cognitive Neighbourhood and Polarization

### 5.4.1 Raising Aspirations: a similarity approach

Assume that the decision-maker is characterized by $p_0$,$\theta_0$ so that she is in a situation of aspirations failure. The question then becomes: how can the aspirations of the decision-maker be raised?

In practice, one way of raising aspirations is by observing role models (Rao and Walton 2004): individuals will draw on the aspirations of their cognitive neighbours, and in this sense role models become an important variable in the formation of their aspirations. But what determines which other individuals are cognitive neighbours of the decision-maker?

Suppose the decision-maker observes an external signal which consists of the initial status of another individual and her achievement $\left(\theta_0^j, \theta_1^j = \Delta' + \theta_0^j\right)$. Under what conditions does the decision-maker update her prior beliefs when she observes an external signal $(\theta_0^j, \theta_1^j)$? In other words, when is such information relevant?

In order to answer this question, following Gilboa and Schmeidler (2001), we endow the decision-maker with a similarity function $s : \Theta \times \Theta \rightarrow [0, 1]$ which provides a quantification of the similarity judgement of the decision-maker, her assessment of how similar the initial status of the role model is relative to the her own initial status. We assume that assessing the similarity across different pairs of initial status is the main cognitive task of the decision-maker. Importantly, the similarity function is subjective in the same sense in which probabilities are subjective in expected utility theory. Gilboa and Schmeidler (2001) provide an axiomatic treatment of choice determined by similarity weighted payoff estimation.

Nevertheless, there may be an objective element in the assessment of similarity. The problem is familiar from econometrics where one might want to infer the conditional distribution $p(y \in A|x_0)$ where the sample frequency of $x_0$ is zero i.e. $p(x_0) = 0$. Assume that all variables are unidimensional. In such scenarios, it is standard in econometrics to use a uniform kernel estimate (Hardle, 1990; Manski, 1999) which is an estimate of the sample frequency with which $y \in A$ amongst those observations $x_i$ such that $|x_i - x_0| < d$ (where $d$ is the sample specific bandwidth chosen to confine attention to those observations in which $x_i$ is close to $x_0$). In a sample with $n$ observations, the expression for the uniform kernel estimate is

$$\frac{\sum_{i=1}^{N} \mathbf{1}(y \in A)\, \mathbf{1}(|x_i - x_0| < d)}{\sum_{i=1}^{N} \mathbf{1}(|x_i - x_0| < d)}. \tag{5.3}$$

Then, the uniform kernel estimate corresponds to a "bandwidth" similarity function where

$$s\left(\theta_0^j, \theta_0\right) = \begin{cases} 1, \text{ if } \left|\theta_0^j - \theta_0\right| \leq d \\ 0, \text{ otherwise} \end{cases} \tag{5.4}$$

A different and continuous choice of a similarity function is

$$s\left(\theta_0^j, \theta_0\right) = 1 - \frac{\left|\theta_0^j - \theta_0\right|}{\left|\overline{\theta} - \underline{\theta}\right|}. \tag{5.5}$$

Consistent with both similarity functions, we assume that $s(\theta_0, \theta_0) = 1$ and that $s\left(\theta_0^j, \theta_0\right)$ is decreasing in the distance (in some metric) between $\theta_0^j$ and $\theta_0$.

Fix the external signal $\left(\theta_0^j, \theta_1^j = \Delta' + \theta_0^j\right)$. Given a similarity function, the decision-maker updates his payoffs from choosing $\overline{a}$ as follows:

$$\begin{aligned} & s\left(\theta_0^j, \theta_0\right)\left[b\left(\Delta' + \theta_0\right) - c\left(\theta_0\right)\right] + p_0 b\left(\Delta' + \theta_0\right) - c\left(\theta_0\right) \\ = \; & \left(s\left(\theta_0^j, \theta_0\right) + p_0\right) b\left(\Delta' + \theta_0\right) - \left(1 + s\left(\theta_0^j, \theta_0\right)\right) c\left(\theta_0\right) \end{aligned}$$

which, after an affine transformation of payoffs, is equivalent to

$$\frac{s\left(\theta_0^j, \theta_0\right) + p_0}{1 + s\left(\theta_0^j, \theta_0\right)} b\left(\Delta' + \theta_0\right) - c\left(\theta_0\right) \tag{5.6}$$

This has the interpretation that after observing the external signal, the decision-maker has updated his prior beliefs so that:

$$p_1 = \frac{s\left(\theta_0^j, \theta_0\right) + p_0}{1 + s\left(\theta_0^j, \theta_0\right)} \tag{5.7}$$

Therefore, the updating of priors after observing the signal (the role model) is an example of similarity based learning.

Remark that in the case when $n = 1$ (the case of a single role model, the case studied so far), $\Theta = \left[\underline{\theta}, \overline{\theta}\right] \subset \Re$ with the interpretation that $x_0$ corresponds to $\theta_0$, $x_i$ corresponds to $\theta_0^j$ and $y \in A$ corresponds to achieving $\Delta'$, the expression for $p_1$ reduces to uniform kernel estimate for the "bandwidth" similarity function.

Therefore, after observing the external signal, the decision-maker will choose $\overline{a}$ iff

$$\frac{s\left(\theta_0^j, \theta_0\right) + p_0}{1 + s\left(\theta_0^j, \theta_0\right)} b\left(\Delta' + \theta_0\right) - c\left(\theta_0\right) \geq 0$$

or equivalently will be a role model iff

$$\frac{s\left(\theta_0^j, \theta_0\right) + p_0}{1 + s\left(\theta_0^j, \theta_0\right)} \geq \hat{p}\left(\Delta, \theta_0\right) \tag{5.8}$$

Observe that if $s\left(\theta_0^j, \theta_0\right) \cong 0$, $p_1 \cong p_0$ and if $s\left(\theta_0^j, \theta_0\right) \cong 1$, $p_1 \cong \frac{1+p_0}{2}$. Moreover, $p_1$ is increasing in $s\left(\theta_0^j, \theta_0\right)$ so that $p_1 \geq p_0$ with the strict inequality whenever $s\left(\theta_0^j, \theta_0\right) > 0$.

So far we have assumed that the external signal takes the form $\left(\theta_0^j, \theta_1^j = \Delta^{j\prime} + \theta_0^j\right)$. However, for the individual who consists of the external signal to choose $\Delta^{j\prime}$, by Proposition 17, it must be the case that

$$p_0^j \geq \hat{p}^j\left(\Delta^{j\prime}, \theta_0^j\right) \text{ and } \Delta^{j\prime} \geq \hat{\Delta}(p_0^j, \theta_0^j). \tag{5.9}$$

where $p_0^j$ denotes the prior probability of the external signal.

Building on the preceding analysis, the following proposition characterizes the conditions under which the external signal will act as a role model:

**Proposition 18** ***(Role model)*** *If* $\Delta' < \hat{\Delta}(p_0, \theta_0)$ *and* $\Delta^{j\prime} \geq \hat{\Delta}(p_0^j, \theta_0^j)$*, the external signal will act as a role model iff* $p_0 < \hat{p}\left(\Delta', \theta_0\right) \leq \frac{s\left(\theta_0^j, \theta_0\right) + p_0}{1 + s\left(\theta_0^j, \theta_0\right)}$.

Proposition 18 states the conditions under which the decision-maker will draw upon the aspirations and achievements of a role model who will alter her choices, aspirations and achievements and show her the way out of the aspirations trap. The key requirement is that it has to be the case that the extrinsic circumstances (the initial status) of the role model has to be similar to the extrinsic circumstances (the initial status) of the decision-maker herself. Thus, the decision-maker will not put much weight on the experience of success of individuals who are characterized by very different to her.

Nevertheless, even if the initial status of the role model were similar that of the decision-maker, the role model will need to have a higher degree of self-confidence.

Therefore, one way to raise the aspirations of all individuals belonging to a disadvantaged group would be to alter the behavior of a carefully chosen subset of such individuals. A different way would be to raise the self-confidence of a disadvantaged individual directly. The two case studies, Classical Music Orchestras for children from disadvantaged backgrounds and the decrease in the HIV infections in Sonagachi (Kolkata's oldest and best established red-light district), discussed in greater detail in Section 6, illustrate the above points.

### 5.4.2 Choosing the Cognitive neighbourhood

It remains to determine the cognitive window of the decision-maker. In determining the cognitive window, note that there is a tension between (a) looking at only those individuals who are similar and (b) observing an individual who has a higher degree of self-confidence and therefore, who has chosen a higher level of aspirations.

For simplicity, we consider a decision-maker such that $p_0 < \hat{p}\left(\Delta', \theta_0\right)$ and $\Delta' < \hat{\Delta}(p_0, \theta_0)$ (for example, assume that $\theta_0 = \underline{\theta}$).

Consider the following decision-process:

Step 1: The decision maker chooses a subset from $\Theta$ sample bandwidth $d \in \left[0, \left|\overline{\theta} - \underline{\theta}\right|\right]$ at a cost $C(d)$ where $d$ is the radius of the subset and $C(.)$ is a strictly increasing, continuous function that captures the idea of costly cognition.

Step 2: Once the decision-maker has chosen $d$, she is able to sample a role model with characteristics $\left(\theta_0^j, \theta_1^j \geq \theta_0^j\right)$ where $\theta_0^j$ is drawn according to the uniform distribution from her chosen subset. After observing the role model, the decision-maker updates her priors using the similarity function $s\left(\theta_0^j, \theta_0\right)$ and re-optmizes.

**Proposition 19** ***(choice of the cognitive neighbourhood)*** *Suppose $\theta_0 = \underline{\theta}$, the similarity function is continuous with $s\left(\overline{\theta}, \underline{\theta}\right) = 0$. Then, the decision maker will choose a cognitive window which is subinterval of $\left[\underline{\theta}, \overline{\theta}\right]$ and $d^*$, the length of the optimally chosen subinterval, characterized by $\int_{\underline{\theta}}^{\underline{\theta}+d^*} p_1\left(\theta_0^j\right)\left(\frac{b\left(\Delta^j+\theta_0\right)}{d^*}\right) d\theta_0^j - c\left(\theta_0\right) \geq C(d^*)$, is strictly less than $\left|\overline{\theta} - \underline{\theta}\right|$.*

**Strongly connected societies and aspirations**

We conclude this section with a simple $N$-person extension that links the nature of inequality in a society with aspirations failure. Consider a society with $N$ (a large but finite number) individuals distributed on $\left[\underline{\theta}, \overline{\theta}\right]$.

**Definition 11** ***(Individual connectedness)*** *Individual $i$ is connected to $j$ if $\theta_0^i < \theta_0^j$ but when $j$ chooses $\bar{a}$, $j$ belongs to the cognitive neighborhood chosen by $i$.*

Define a directed graph over $N$ where the vertices are individuals and an arc $(i, j)$ exists iff $i$ is connected to $j$. A path in a directed graph is an ordered collection of arcs and vertices in which all vertices are distinct. A directed graph is strongly connected if for every pair of distinct vertices $(i, j)$ there exists a path connecting $i$ to $j$.

**Definition 12** ***(Strongly connected society)*** *A society is strongly connected if its associated graph is strongly connected.*

The distribution of individuals in a society is $\varepsilon-$dense if for $\varepsilon > 0$, there exists a pair $i, j$ with $\left|\theta_0^i - \theta_0^j\right| < \varepsilon$.

**Proposition 20** ***(Connectedness and aspirations)*** *There exists $\bar{\varepsilon} > 0$ such that if the distribution of individuals in a society is $\varepsilon-$dense, $0 < \varepsilon < \bar{\varepsilon}$, the society is strongly connected and all individuals in it choose $\overline{a}$ and have aspirations $\Delta'$ as long as there exists one individual $k$ with $p_0^k \geq \hat{p}\left(\Delta^{k\prime}, \theta_0^k\right)$. More generally, the probability that all individuals in it choose $\overline{a}$ and have aspirations $\Delta'$ is $\max_{k\in N} 1 - F(\hat{p}\left(\Delta^{k\prime}, \theta_0^k\right))$.*

The following example examines what happens when the distribution of individuals fails to be sufficiently dispersed.

**Example 17** *Polarization and aspirations failure*

Consider a society where a fraction $\alpha$ individuals are initially located at $\underline{\theta}$ and $1-\alpha$ individuals are initially located at $\overline{\theta}$ with $s\left(\overline{\theta},\underline{\theta}\right)=0$. Assume that $\hat{p}\left(\Delta',\underline{\theta}\right)>1$ but $\hat{p}\left(\Delta',\overline{\theta}\right)\approx 0$.

Then, clearly $p_0<\hat{p}\left(\Delta',\theta_0\right)$ and $\Delta'<\hat{\Delta}(p_0,\theta_0)$ for all $\theta_0=\underline{\theta}$ while $p_0'\geq\hat{p}\left(\Delta',\theta_0'\right)$ and $\Delta'\geq\hat{\Delta}(p_0',\theta_0')$ for all $\theta_0'=\overline{\theta}$ and no $\theta_0'=\overline{\theta}$ is in the cognitive neighborhood of any $\theta_0=\underline{\theta}$. Aspirations failure for all individuals initially located at $\underline{\theta}$.

The next example examines what happens if the distribution of individuals is concentrated at a specific initial status.

**Example 18** *Equality and aspirations failure*

Consider a society where all individuals are initially located at $\underline{\theta}$. Assume that $\hat{p}\left(\Delta',\underline{\theta}\right)>1$. Then, clearly $p_0<\hat{p}\left(\Delta',\theta_0\right)$ and $\Delta'<\hat{\Delta}(p_0,\theta_0)$ for all $\theta_0=\underline{\theta}$ and there is aspirations failure for all individuals initially located at $\underline{\theta}$. More generally, the probability that all individuals choose $a=\overline{a}$ and $\Delta=\Delta'$ is $1-\left[F(p\left(\Delta',\theta_0\right))\right]^N$.

## 5.5 Preference formation and aspirations failure

Assume that there is some underlying preference formation process as in BDPs $\pi:\{\underline{a},\overline{a}\}\times[0,1]\times\Theta\rightarrow[0,K]$ from actions, self-confidence and initial social status to aspirations. Assume that $\pi(a,p,\theta)$ is increasing in $a$ and $\theta$ with $\pi(\underline{a},p,\theta_0)=0$ for all $\theta_0\in\Theta$ and $\pi(\overline{a},p,\theta_0)=\overline{\Delta}\left(p,\theta_0\right)\in(0,K]$ where $\overline{\Delta}\left(p,\theta_0\right)$ is increasing in $p,\theta_0$: therefore, we assume that aspirations can be constrained by both actions and initial status.

We can think of two different ways in which the aspirations of the decision-maker can be endogenized.

First, the decision-maker internalizes the preference formation process: let us label such a decision-problem as a standard decision problem. At this point, there are two cases to consider. In the first instance, decision-maker chooses a $\Delta$ from $[0, K]$ taking into account the fact that by changing $\Delta$ she affects her own choice. Alternatively, the decision-maker chooses actions taking the feedback into account. In the first case, by Proposition 15, the decision-maker will anticipate that she will choose $\overline{a}$ iff $\Delta' \geq \hat{\Delta}(p, \theta_0) > 0$. In the second case, if the decision-maker chooses $\overline{a}$, she anticipates that the feedback generates an aspiration level of $\overline{\Delta}(\theta_0)$. Therefore, the decision-maker will choose $\overline{\Delta}(\theta_0)$ (equivalently, $\overline{a}$) iff

$$pb(\overline{\Delta}(\theta_0) + \theta_0) - \overline{c}(\theta_0) \geq 0 \tag{5.10}$$

$\Leftrightarrow$

$$\overline{\Delta}(\theta_0) \geq \hat{\Delta}(p, \theta_0). \tag{5.11}$$

Second, the decision maker takes the aspirations as given while choosing action although both aspirations and actions are required to be mutually consistent given the preference formation process. This is an application of the BDP introduced in Chapter 2. In this case, if

(i) $\overline{\Delta}(p, \theta_0) < \hat{\Delta}(p, \theta_0)$, the unique outcome is $(a = \underline{a}, \Delta = 0)$;

(ii) if $\overline{\Delta}(p, \theta_0) \geq \hat{\Delta}(p, \theta_0)$ and $\overline{\Delta}(p, \theta_0) < K$ there are two outcomes, $(a = \underline{a}, \Delta = 0)$ and $\left(a = \overline{a}, \Delta = \overline{\Delta}(\theta_0)\right)$;

(iii) if $\hat{\Delta}(p, \theta_0) \leq \overline{\Delta}(p, \theta_0) = K$, the unique outcome is $\left(a = \overline{a}, \Delta = \overline{\Delta}(p, \theta_0)\right)$.

Call $(a = \underline{a}, \Delta = 0)$ a type I outcome and $\left(a = \overline{a}, \Delta = \overline{\Delta}(p, \theta_0)\right)$ a type II outcome.A type I outcome can be interpreted as an aspirations failure, a low motivation trap for the individual.

The set of equilibria is "weakly increasing" in $\theta$. For an individual of low initial status (low $\theta_0$), the unique equilibrium is type I while for an individual with high initial status (high $\theta_0$) the unique equilibrium is type II.

For an individual in the middle, with intermediate values of $\theta$, there are multiple welfare ranked equilibria and for such an individual, the theoretical framework developed so far doesn't pin down the equilibrium decision state i.e. the aspirations and choices are indeterminate.

Formally, the set of low initial status is $\underline{\Theta} = \left\{\theta_0 : \overline{\Delta}\left(p, \theta_0\right) < \hat{\Delta}(p, \theta_0)\right\}$, the set of high initial status is $\overline{\Theta} = \left\{\theta_0 : \hat{\Delta}(p, \theta_0) \leq \overline{\Delta}\left(p, \theta_0\right) = K\right\}$, and the set of intermediate status is $\Theta_M = \left\{\theta_0 : \overline{\Delta}\left(p, \theta_0\right) \geq \hat{\Delta}(p, \theta_0), \overline{\Delta}\left(p, \theta_0\right) < K\right\}$. Assume that all the three sets $\underline{\Theta}, \overline{\Theta}$ and $\Theta_M$ are non-null and constitute a partition of $\Theta$.

Next, we develop a simple adaptive dynamics over $\Delta$ that allows us to select between the different outcomes for $\theta_0 \in \Theta_M$. Fix $\theta_0 \in \Theta_M$ and consider the following adaptive dynamics over $\Delta$:

Step 1: The initial $\Delta$ is picked at random from $[0, K]$ (where $K$ is some large but finite number) according to some continuous pdf $g(p)$ (with associated cdf $G(p)$)

Step 2: Given the $p$,$\Delta, \theta_0$, the individual chooses $a \in \{\underline{a}, \overline{a}\}$. The new status $\theta_1$ is generated by the production function where $\Delta = \pi\left(a, p, \theta\right)$.

The following proposition formalizes the link between aspirations and achievement and initial disadvantage in a behavioral decision problem.

**Proposition 21** ***(Disadvantage and aspirations failures in BDPs)*** *There exists $\tilde{\theta}$ such that if $\theta_0 < \tilde{\theta}$, a type I outcome will have a higher probability of emerging while if $\theta_0 > \tilde{\theta}$, a type II outcome will have a higher probability of emerging.*

## 5.6 Policy discussion

If economic betterment is an important goal, in addition to redistributive measures, the aspiration neighborhood of individuals also need to be improved. Thus, policy intervention should provide the mechanisms to help the individual to change her aspirations. These mechanisms are extremely sensitive to apparently irrelevant “details” from the point of view of someone who has not experienced this condition (i.e. the policy maker), but that are very relevant for people caught within an aspiration trap.

For example, the importance of role models (Rao and Walton, 2004) cannot be understood within the framework of a standard cost-benefit analysis. Individuals will draw on the aspirations of their cognitive neighbors, and in this sense role models become an important variable in the formation of their aspirations.

Changing a role model to break an aspiration trap has been empirically showed to be a very effective policy of poverty reduction. One of the most remarkable examples comes from Kolkata (Calcutta), India. The objective of the Government by the 1990s was to decrease the HIV infections in Sonagachi, Kolkata’s oldest and best established red-light district, with over 4,000 sex workers working in 370 brothels that service about 20,000 clients a day (Rao and Walton, 2004, pg. 7). As Rao and Walton (op. cit.) points out, during the 1990s, the Government’s "interventions tended to reflect the values of the middle-class bureaucrats who crafted them. They focused on rehabilitating the sex workers, rescuing them, and [...] training them." to be ready for an insertion in the "good life". This strategy did not work. As Rao and Walton argues, the relatively high earnings in sex work and the discrimination faced by former sex workers in the world outside Sonagachi, led most women to return prostitution.

In this context, the Government implemented a very different strategy. It decides to convince and train twelve sex workers which would pass the important information about the use of condoms as peers education to their co-workers. This process, as argued by Rao and Walton, led over a period of two or three years, to a "metamorphosis" in the sex worker's aspirations. Moreover, the program was remarkably successful as a health intervention, with almost all sex workers using condoms at least some of the time and the HIV incidence in Sonagachi substantially decreased.

This successful health intervention can be interpreted with the lens of our model as a change in the quality of the aspiration windows. The individuals incorporated as new aspiration neighbors their renewed peers through a re-identification process.

In addition, the results in this chapter offer new insights about why programs that look for increasing the aspirations of the people by means of improving their aspiration neighborhood such us the Venezuelan Classical Music Orchestras for children from disadvantaged backgrounds or role models programs are so successful. More than 400,000 Venezuelan children, most coming from poor families have taken and take part of a network of orchestras directed by Jose Antonio Abreu that began being carried out 30 years ago. The lessons are free of charge and a public foundation "FESNOJIV", which is also supported by the Inter American Development Bank, provides the instruments. The project does not primarily aim to create professional musicians, but to integrate poor children into the society. The recognition children get from those who are out of their aspiration neighborhood makes them change their beliefs about themselves. Music becomes the driver that makes the social integration of different Venezuelan population groups possible. 96 percent of the young musicians have good to excellent school records. They stand out as high achievers thanks to their steady relationship with music. UNESCO awarded FESNOJIV its International Music Award in 1993-94 and in 1998 UNDP commended it as an outstanding example of poverty reduction. In the words of its founder: "The majority of the children and juveniles belong to the groups that are most vulnerable and excluded in all of Venezuelan society. Participating in the orchestral movement has made it possible for them to set up new goals, plans, projects and dreams, and at the same time it is a way of creating meaning and helping them in their day-to-day struggle for better conditions of life through the variety of opportunities that the orchestral movement offers them." Antonio Abreu (see http://www.rightlivelihood.org/recip/abreu.htm)

More generally, structural poverty reduction must consider empowering of individuals. Our results can provide one possible explanation of why some paternalistic policies have failed (Narayan, 2002) and why in general, externally applied incentives lead to "passive acceptance at best and more likely to indifference or resistance's" (McGregor, 1960, pg. 68).

Regarding the potential of our model for the impact on poverty reduction agenda, our model would suggest that a policy that aims to succeed reducing structural poverty requires additional reinforcement mechanisms to be at work: such a policy should pay special attention to be inclusive and empower the poor. As Ellerman (2004, pg. 110) argues, "reliance on carrots and sticks can induce an atrophy effect. Any original intrinsic motivation dries up, and the doer becomes an aid-dependent marionette responding only to external strings."

In a related analysis, Bertrand et al. (2004) and Mullainathan (2005) point out that "small" institutional barriers that would appear insignificant in a cost benefit analysis would become very psychologically costly for poor people within an aspiration trap. "Small" hassles can explain low welfare program take-up. Application processes often signals negative identities that can reinforce the alienation and lack of aspirations of the eventual program taker. The importance of such hassles starts to become clear in the analysis proposed here.

In a similar vein, major policy interventions that do not consider the importance of psycho-social constraints such aspirations may fail to break poverty traps. For instance, creating jobs is not necessarily an effective policy to solve an aspiration failure. As Atkinson (1998) argues, ending persistent poverty will depend on the nature of these new jobs. Do they restore a sense of control? Do they offer prospects for the future? Likewise, income support programs may be exclusionary, as they could make the recipients feel excluded by the state.

## 5.7 Conclusion

In this chapter we proposed a model that associates people initial disadvantages with their self-confidence, aspirations and decisions to change their status-quo. Our analysis suggested a simple view of a new class of poverty traps that does not rely only on external constraints to the individual, but also on internal psychological constraints. We found that initial disadvantages are associated with aspirations failures, a combination of low aspirations and low achievement. We then studied how and under what conditions a role model can help a poor person to break her aspirations failure and climb out of poverty. We allowed the person to select a role model and showed that, in principle, a poor person may restrict the set from which the role model is selected. Evidence from Anthropology, Sociology and Social Psychology supports our results and our model. The fact that the actual condition of poverty affects (and is affected by) lack of aspirations and hope is a recurrent topic in these disciplines. However, it is surprising to see that it has been disregarded by formal economic models. The concept of empowerment has also been understudied in Economic Theory, despite its great importance for the anti-poverty policy agenda.

Our preliminary results are consistent with empirical evidence suggesting a tendency of societies to polarize(see Duclos et al., 2004; Quah, 1993; Azariadis and Stachurski, 2005; Atkinson, 2003). Similarly, our results are in line with empirical evidence showing that communities with higher levels of social cohesion and narrower gaps between poor and rich produce better health and welfare outcomes than wealthier societies that have higher levels of social disintegration (Putnam, 2000). Marmot and Wilkinson (2006) show that in addition to economic prosperity, equality and social cohesion are also powerful determinants of health. The case of Kerala provides further evidence suggesting that the social cohesion gained by allowing the poor to participate in programs contributes to the achievement of high positive indexes of Human Development (see Kannan, 2000)

# Chapter 6

# Aspirations and Human Cooperation: Experimental Evidence

## 6.1 Introduction

Aspirations are commonly defined by psychologists as goals, which can be either intrinsic (e.g. meaningful relationships, personal growth, community contributions) or extrinsic (e.g. annual income, wealth, fame, image) and long term (e.g. win the Nobel Prize) or short term (e.g. get a degree, finish a paper).

While standard economics analysis focuses on individuals maximizing utility, the concept of individuals satisfying aspirations has been shown to be central in understanding phenomena in Social Psychology (Deci and Ryan, 2000), Sociology (Sewell et al., 1957; Sewell, 1964; Sewell and Shah, 1968), Antropology (Appadurai, 2004) and decision theory (Simon, 1955, 1959; Selten, 1998)[1]. It is used to explain such phenomena as educational and occupational choice (Sewell, 1964), the Easterlin paradox (Easterlin,

[1]See Traub (1999) for a review of aspirations as frames that individuals use to make decisions.

2001), mental health (Kasser and Ryan, 1993), mental wellbeing (Ryan et al., 1999) and, more recently, poverty persistence (Ray, 2006).

By comparison, human cooperation vs. free-riding in a context of private provision of a public good has always been a challenge to social scientists (Dawes and Thaler, 1988). With the globalization process, this challenge has become even more complex, since an additional geographical dimension was added to the problem: the global dimension of public goods. The challenge now is not only to understand whether people contribute or not to a public good, but also if the degree of cooperation change which the geographical level of the public good (local, national or global).

This chapter provides a first attempt to explore (a) the link between these two central concepts, human aspirations and cooperation, and (b) if this link changes depending on whether the contributions are directed to a local, national or public good. Specifically, it enquires into what extent people's real life income aspirations affect their level of cooperation in general and, in particular, with whom they cooperate.

In order to grasp the intuition behind the aspirations-cooperation link, let me give you the following example. Susan is working to submit a paper for an important conference tomorrow. This is her personal goal. At the same time, she has also committed herself to spend the day cleaning a common area together with her neighbour. Susan cannot do both task, so she has to choose one. Suppose that Susan individually prefers submitting her paper to cooperating with neighbours, while having the common area cleaned is unanimously preferred to leaving it as it is. Suppose also Susan is always better off when some neighbour cleans than when none does. Will she cooperate with her neighbours and leave her individual goal of submitting the paper unattained or will she simply follow her own interest? This example is a version of an n-person prisoners' dilemma, and the answer to the question is trivial in standard Economics: Susan is rational and selfish, and assuming no punishment, she will defect. Recent literature on social preferences allows for the possibility of assuming that Susan may also care about the welfare of her neighbours (either unconditionally altruistic or by warm glow effect) and so may leave her goal unattained and cooperate. She might also care about reciprocity, and if she thinks their neighbours have been good to her, again she may be willing to cooperate with them. Now consider a different scenario. Say she is at a point in which she is somehow satisfied with what she has written, although if she continues polishing the paper, it can only be improved further. Would she help her neighbours in this second scenario? Standard Economics would predict that, again, a selfish and rational Susan will free-ride on her neighbour in this second scenario. There is neither an existing Economic Theory nor empirical evidence that would support a different answer. The trade-off between Susan's and the neighbours' welfare is assumed to be fixed, and it is independent on Susan having reached or being close to reaching a satisfaction/aspiration target. This chapter suggests, however, the existence of such a target, and claims that Susan would be more likely to cooperate with her neighbour in this latter scenario than in the former. In addition, the chapter studies whether Susan's decision in each scenario would be different if instead of having to clean a common local area, she had to help Greenpeace to clean the river that would benefit citizens of a border country (global public good) instead of her neighbours (local public good).

In order to explore these concerns, this chapter combines experimental data with a post-experiment survey. The experimental design was constructed by Buchan et al. (2009) and it consists in a nested PGG similar to that employed by Blackwell and McKee (2003) in Economics and by Wit and Kerr (2002) in Psychology. In a nested PGG, individuals have the option of keeping their endowment for themselves, contributing some of it to a local pot (L), and/or contributing some of it to a larger pot – in this case representing either a national (N) or global pot (G). In our experiment, participants play two interactions involving this nested PGG, one in which the contribution entails a choice between the local and national pot (Decision N), and a second in which the contribution entails a choice between the local and global pots (Decision W). The local pot is comprised of the participant plus three other participants from the local area. The national or global pot consists of the participant's local group plus two other groups of four people from different areas of the same country (Decision N), or from different countries around the world (Decision W). A post-experiment survey collects different variables of interest for this chapter, such as social identity measures, demographic measures and measures of income aspirations gap, which is defined as the proportion of income needed by the participant in order to be satisfied income-wise.

We are interested in several questions. First, are participants who are more satisfied income-wise more likely to provide higher contributions to a public good than those who are less satisfied? If so, to what extent? Does this behaviour depend on whether the public good is local, national or global? Second, will the in-group identification with neighbours, compatriots or foreigners affect the level of contribution to the local, national and global public good? Third, will the perceptions of participants' position in the income distribution with respect to the average neighbour, compatriot or foreigner affect their contributions to each public good? Will this perception affect the way in which the income aspirations gap affect their contributions?

The rest of the chapter is organized as follows. Section 6.2 presents the methodology, describes the data and defines the hypotheses to test. Section6.3 presents the analysis and main results. Section 6.4 discusses the robustness of the results and Section 6.5 concludes.

## 6.2 Method

### 6.2.1 Experiment and Survey

Data from a linear public-good experiment was gathered and a survey conducted in Argentina. The experiment was part of a set of similar experiments conducted in six other countries around the world[2] under the umbrella of a meta-project on cooperation and

[2] The experiment was conducted also in Iran, England, Italy, United States, South Africa and Russia.

globalization (Buchan et al., 2009). I only make use of the data gathered in Argentina, where I included additional relevant questions for this chapter.

Approximately 200 subjects participated in three public goods experimental decisions[3] in fixed order. Pilot tests found no ordering effects. Subjects were drawn from the general population in Buenos Aires (city and province)[4]. A quota sampling recruitment method was used based on gender, age and socioeconomic status[5]. The experimental sessions were conducted in groups of no less than eight and no more than eighteen participants. Interactions were anonymous, and the groups with which subjects interacted were randomly selected at the beginning of each decision. No feedback between decisions was provided. Hence, the three decisions can be treated as independent.

In each of the two decisions studied here, subjects were given 10 tokens each[6]. The task in each decision was to decide how to allocate these 10 tokens among their "personal" account (or private good) and two "collective" accounts (or public good), whose composition varied across the two decisions. Each token put into the "personal" account maintained intact its value (i.e. the individual marginal return to the *private* good, $\alpha$, was 1). As it is standard in LPGGs, the individual marginal return to the *public good*, $\beta$, is lower than the marginal return to the private good, $\alpha$, but contributing creates positive externalities for the other people in the group (i.e. $\alpha > \beta$, $N\beta > \alpha$, where $N$ is the number of players).

In a first decision, Decision N, subjects were asked to allocate their endowments between their "personal" account, a "local" account (or local public good) composed of the subject and another three anonymous randomly chosen neighbours, and a "national"

[3]Participants' play three Linear Public Good Games (LPGG) or Voluntary Contributions to a Public Good Game. For the purpose of this paper, we are going to focus only on two of the three decisions played, simply because this two decisions have exactly the same structure of incentives and only differ in the identity of the recipients of the public good.

[4]The locations chosen were Almagro, Boedo, Caseros, San Isidro and Las Flores. Approximately 50 subjects were recruited per location. Subjects recruitment was carried out by the CINEA, an argentine agency specialised in survey polls and market research.

[5]Gender (male, female), age (18-30, 31-50, 51-70) and socio-economic status (low, medium, high). The administration of the experiment was oral and paper-based.

[6]One token was worth the purchasing power equivalent of US $0.50

account (or national public good). The "national" group was made up of the subject, the same three "local" people benefiting from the "local" account - plus eight anonymous subjects from other parts of Argentina. Tokens allocated to the "personal" account were saved for the individual ($\alpha = 1$). Tokens placed in the "local" account were summed up, and the total was doubled and shared equally between the four individuals of the "local" group. Tokens placed in the "national" account were instead tripled by the experimenter and divided equally among the 12 people of the "national" group. Thus, the marginal return of one token allocated to the "local" account was $\beta^L = 0.5 = \frac{(1\times2)}{4}$, and to the "national" account was $\beta^N = 0.25 = \frac{(1\times3)}{12}$. Therefore, the payoff for individual $i$ in decision N was:

$$\pi_i^N(x_i, G) = \alpha x_i + \beta G = \alpha(10 - g_i^{L+N}) + \beta^L(G_{-i}^L + g_i^L) + \beta^N(G_{-i}^N + g_i^N), \qquad (6.1)$$

where $x_i \in [0, 10]$ is the number of tokens that player $i$ keeps for herself (private good), $g_i^k \in [0, 10]$ for $k = \{N, L\}$ is the number of tokens player $i$ contributes to the $k$ collective account (or public good) and $G_{-i}^L = \sum_{j=1}^{N} g_j^k$ for $j \neq i$ is the sum of the contributions of the other members of the group. $x_i$ and $g_i^k$ are positive integers.

In a second independent decision, Decision W, subjects chose how much to allocate among their "personal" account, the "local" account and the "world" account. The structure of incentives in Decision W was exactly the same as that in Decision N (i.e. $\alpha = 1 > \beta^L = 0.5 > \beta^N = \beta^W = 0.25$) and **only** the composition of the group differed. The "world" group was made up of the subject, three "local" people - plus eight anonymous subjects from different countries around the "world." Subjects were not told which countries these other subjects were from, but were informed that these countries could be from any of the four continents where the research was conducted[7].

[7]Due to the logistic of research, it was impossible to run the experiments simultaneously within a single country. Therefore, the team of scholars who designed the macro experiment had to rely on a 'dynamic' matching procedure, where past decisions from other participants were used to determine

Therefore, the payoff for individual $i$ in decision W was:

$$\pi_i^W(x_i, G) = \alpha x_i + \beta G = \alpha(10 - g_i^{L+W}) + \beta^L(G_{-i}^L + g_i^L) + \beta^W(G_{-i}^W + g_i^W), \tag{6.2}$$

Both decisions, N and W, characterize a multilevel public good dilemma. It is individually optimal to contribute nothing to the "collective" accounts ($\alpha > \beta^k, k = \{N, L, W\}$) although this goes against the social welfare of the group. In other words, the Nash equilibrium of this game is Pareto sub-optimal. These two decisions aim to mimic a situation in which the person has to decide wether to contribute to a *local public good* and/or *national public good* (decision N) and to a *local* and/or *global public good* (decision W). The structure of incentives in both decisions is identical, the only change being the nationality of the co-members of the "collective" accounts. This feature allows us to identify, among other things, how the identity of the people eventually affects people's cooperation directly or/and mediates the effect of individual aspirations on the propensity to cooperate. In order to maximize the validity of our results, we controlled for the subjects understanding of the incentives and rules of the game[8] and experimenter effects[9].

### 6.2.2 Data Description

The dependent variable, "propensity to cooperate," is measured by individual contributions to the public goods, i.e. to the "collective" accounts. The independent variables

the payoffs of subjects currently taking part in the research. When possible, matching at the local level happened among people taking part in the same session. An algorithm was produced that ensured that a subject's choices entered the dataset as the experiments ensued. These decisions were used to match people's decisions in subsequent sessions. Starting data for these decisions was provided by pilot tests which occurred in each country prior to the experiment and through a series of pilot tests conducted in four countries during the prior two years. For more information about this and other particularities of the macro experiment, please refer to Buchan et al. (2009).

[8] Built into the experiment instructions there was a basic understanding check, in which subjects were asked to answer some questions about the interaction. This gave us the chance to check for subjects' comprehension of the task. Subjects (6 out of 207 subjects) who showed evident failures to understand the task have been expunged from the dataset.

[9] A detailed protocol was followed by the experimentet in all the sessions.

used in this analysis come from an individual survey that participants completed at the end of the experiment.

### Observed behaviour

Table I provides descriptive statistics for the dependent variables.

TABLE I: DESCRIPTIVE STATISTICS OF DEPENDENT VARIABLES[10]

| | mean | sd. | max. | min. | N |
|---|---|---|---|---|---|
| Total Cont. (N) | 7.32 | 2.67 | 10.00 | 0.00 | 201 |
| Cont. "local" Account (N) | 3.10 | 2.01 | 10.00 | 0.00 | 201 |
| Cont. "national" Account | 4.22 | 2.65 | 10.00 | 0.00 | 201 |
| Total Cont. (W) | 7.02 | 2.72 | 10.00 | 0.00 | 201 |
| Cont. "local" Account (W) | 3.22 | 2.37 | 10.00 | 0.00 | 201 |
| Cont. "world" Account | 3.81 | 2.84 | 10.00 | 0.00 | 201 |

**Observation 1: high level of contributions**. On average, participants contributed at very high levels in both decisions. More than 70% (in decisions N and W) of people's endowments were allocated to their "collective" accounts.

**Observation 2: more to the nation**. Even though the material incentives in Decision N and W are exactly the same, contributions to the "world" account were significantly lower[11] than contributions to the "national" account[12].

Table II provides the frequency of the tokens contributed in different decisions.

**Observation 3: few free-riders.** Only a very small proportion of participants (6%) contributed nothing at least once to the "collective" accounts, whereas more than 25%

[10] *Note*: Total Cont. N and W stands for Total contributions to the collective accounts in Decisions N and W, respectively.

[11] $t = 1.5231, Pr(T > t) = 0.0643$

[12] However, as expected, no significant difference between the average contribution in Decision N and W was found ($t = -1.1097$, $Pr(|T| > |t|) = 0.2678$).

of the people contributed all their tokens to some "collective" account, in at least one decision. Moreover, only 3.5% of the participants always played their dominant strategy, i.e. did not contribute a single token in any decision, while 30 participants allocated all their endowments in the "collective" accounts in all the decisions. The distribution of free-riders across locations was even.

**Observation 4: interior solutions.** 65% (dec. N) and 69% (dec. W) of the participants played interior solutions, i.e. they split their endowments. This behaviour cannot be predicted with standard game theory in which only material payoffs matter, since corner solutions are the only theoretical predictions of a Linear Public Good Game.

**Observation 5: preference reversals.** More than 60% of the participants changed their contributions in both decisions, although the material incentives were exactly the same.

TABLE II: FREQUENCY OF OVERALL CONTRIBUTIONS

| Tokens contributed | Decision N | Decision W |
|---|---|---|
| 0 | 5% | 6% |
| 1 | 0% | 0% |
| 2 | 0% | 0% |
| 3 | 2% | 3% |
| 4 | 5% | 4% |
| 5 | 5% | 7% |
| 6 | 11% | 13% |
| 7 | 14% | 17% |
| 8 | 20% | 17% |
| 9 | 6% | 6% |
| 10 | 30% | 25% |

### Independent Variables

Table III (see appendix) shows descriptive statistics for the independent variables used in this analysis. Below there is a description of how each independent variable is measured.

Income is measured with the following question:

> "Below you will find a scale of monthly income[13]:
>
> | $5-$350 | $350-$500 | $500-$700 | $700-$850 | $851-$1050 | $1051-$1300 | $1300-$1620 | $1621-$2100 | $2100-$3000 | $3000-$108000 |
> |---|---|---|---|---|---|---|---|---|---|
> | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
>
> I would like to know which bracket your family belongs to, including wages, salaries, pensions, and other income. Check the income bracket that corresponds to your family, before taxes and other deductions."

Following Ray (2006), the variable *Income Aspirations Gap* - denoted by "$g_i$"- is defined as the relative difference between the standard of living that is aspired to - measured by the variable *Absolute Aspirations Level* and denoted by "$a_i$"- and the standard of living subjects already have - measured by the variable *Income* denoted by "$s_i$".- Hence, the *income aspirations gap* is:

$$g_i(a_i, s_i) \equiv \frac{a_i - s_i}{a_i} \tag{6.3}$$

Information about *Absolute Aspirations Level,* $a_i$, is gathered by asking subjects the following question:

> "Consider the following scale of monthly income, [which is the same as the one above] how much monthly income would your household need to be satisfied? Check the income bracket that corresponds to that level of income."

[13] This scale corresponds to the Argentine total household monthly income distribution by deciles at the moment the experiment was designed (July 2006).

Thus, the maximum level of income aspirations gap that a person can have is $\bar{g}_i(a_i, s_i) \equiv \frac{10-1}{10} = 0.9$ and the minimum level is $\underline{g_i}(a_i, s_i) \equiv \frac{1-10}{1} = -9$. I allow for negative income aspirations gap because, although highly unlikely, it is possible that people may be satisfied with less income than that which they actually have. This was indeed the case with some participants distributed across the four locations[14]. A person characterized by $\bar{g}_i(a_i, s_i)$ is said to be "fully gapped": she belongs to the $1st$ decile of the population but aspires to be in the $10th$ decile. On the contrary, someone with a $g_i(a_i, s_i) \leq 0$ is said to be "fully satisfied" income-wise. We interpret $g_i(a_i, s_i) > 0$ as a measure of how frustrated the person is with the income of his household.

During the analysis, other sets of variables were used to capture different aspects that may either affect individual propensity to cooperate, or mediate/moderate the influence of income aspirations gap on it.

Income aspirations gaps may be highly correlated with life satisfaction. In order to eliminate this source of omitted variable bias, participants were asked the standard question used in the subjective well-being literature:

> "Overall, how satisfied are you with your life?[15]

Likewise, the motivations that people may have to cooperate with their compatriots or neighbours may be different from the motivations to cooperate with a foreigner. I collected information on individual degree of identification (social identity) with people from their neighbourhood, Argentina and the rest of the "world" (Yuki et al., 2005, see), with questions such us:

> "to what degree do you feel committed to your "neighbourhood, "Argentina" and the "rest of the world"?; to what degree do you see yourself

[14] The minimum $g$ of our database was $-1$, and it is highly unlikely to have cases of aspirations gap less than this value.

[15] (1=very unsatisfied; 2=somewhat satisfied; 3=somewhat unsatisfied; 4=very satisfied)

as part of your "neighbourhood," "Argentina" and the "rest of the world"?, how close do you feel to the other members of ...?"[16].

In addition, a question aiming to get an approximate measure of people's perceptions of their own income relative to the income of the average recipient of the public good was asked. The question was as follows:

> "could you please indicate in which position you think the income of your household is with respect to the average income of a household of "your neighbourhood", "Argentina" and the "rest of the world".[17]

Finally, questions on *Gender,* (1=Masculine or 2=Feminine) and *Education* (1=Elementary School, 2=High School, 3=Tertiary, 4=University, 5=Masters, 6-Ph.D.) were included in the survey.

### 6.2.3 Estimation

Ordered Logit estimations were used to estimate whether - and eventually how - people's income aspirations gaps affect their propensity to cooperate, ceteris paribus[18]. An ordinal regression model was chosen to allow the distances between categories (i.e. numbers of tokens allocated to different accounts) to differ[19]. One single regression for each decision was carried out. The underlying latent variable was the individual propensity to cooperate with each of the three groups: "local," "national" and "global" ($k = L, N, W$).

[16] Responses were indicated on 4-point scale from "not at all " to "very or a lot."

[17] The options were: 1- Very below average, 2- fairly below average, 3- aprox. same as average, 4- fairly above average, 5- very above average.

[18] I discuss the issue of causality in the last section of this chapter.

[19] I performed a likelihood ratio (LR) test on all the regressions in this paper to test the proportional odds assumption. It turns out that in most of the cases, we cannot reject the null hypothesis that that there is no difference in the coefficients. So, given that in almost all the cases the proportional odds assumption holds, for simplicity we preferred to present the results using a standard rather a generalized ordinal regression model.

The model estimated is as follows:

$$g_i^{k*} = \beta_1 gap_i + \beta_2 S_i + \beta_3 R_i + \beta_4 (gap_i * R_i) + x_i' \beta + \varepsilon_i \tag{6.4}$$

where $gap_i$ is income aspirations gaps, $S_i$ is social identity, $R_i$ is perceptions of relative income, $x_i$ is a vector of control variables including demographic variables, absolute income, and absolute aspirations, social capital, among others. $g_i^{k*}$ is a continuos unobserved number form 0 to 10 representing the propensity to contribute to $k$ account. It is measured in 10 "m" discrete intervals by the number of tokens each people allocate to each $k$ account. It is assumed that

$$g_i^k = m \quad if \quad \tau_{m-1} \leq g_i^{k*} < \tau_m, \quad \text{for } m = 0, ..., 10$$

The probability of contributing $g_i = m$ number of tokens to the $k = \{L, N, W\}$ account was estimated for a given vector of exogenous variables $\mathbf{x}$ and assuming that $\varepsilon_i$ follows a logistic distribution.

$$\Pr(g_i^k = m \big| \mathbf{x}) = F(\tau_m - \mathbf{x}\boldsymbol{\beta}) - F(\tau_{m-1} - \mathbf{x}\boldsymbol{\beta}), \text{ where } F \text{ is the cdf for } \varepsilon \tag{6.5}$$

for $g_i = 0$, $F(-\infty - \mathbf{x}\boldsymbol{\beta}) = 0$, and for $g_i = 10$, $F(\infty - \mathbf{x}\boldsymbol{\beta}) = 1$

### 6.2.4 Hypotheses

The four hypotheses tested in this chapter are the following.

**H.1(Aspirations)** $\beta_1 < 0$. Income Aspirations Gap is negatively correlated with people's propensity to cooperate. In other words, it is expected that, in average, the narrower the income aspirations gap, the higher the number of tokens contributed, ceteris paribus.

The intuition behind this hypothesis is clear: the marginal rate of substitution between individual and collective welfare is decreasing on the extent to which individual's income aspiration is satisfied. Importantly, note that this statement is different to saying that individual and collective welfare are somehow complementary. Our statement implies a one-way causality consistent with some implicit quasi-lexicographic order between the two dimensions (individual welfare and others' welfare). We want to investigate whether people care first about satisfying some material needs or aspirations, and as long as those needs are relatively satisfied, they start caring about the welfare or needs of the others. Implicit in this statement, it is the view that people may not be intrinsically either selfish or altruist as the existing literature seems to suggest, but their level of cooperation may depend on how satisfied their material aspirations are. This hypothesis is grounded on two orthogonal and well-known old theories. First, it relies on Simon's (1955) idea that people look for achieving satisfactory targets instead of maximizing utility. Simon (1955) illustrates his idea with an example of house prices.

> "[An individual selling his house] may regard $15,000 as an "acceptable" price, anything over this amount as "satisfactory," anything less as "unsatisfactory." In psychological theory we would fix the boundary at the "aspiration level"; in economic theory we would fix the boundary at the price which evokes indifference between selling and not selling (an opportunity cost concept)." (pg. 104)

In addition, the hypothesis uses the theory of Hierarchy of Needs developed by Maslow (1943). Individual material aspirations correspond to the lower steps of Maslow's pyramid and social welfare can be understood as the upper levels of the pyramid. The two theories support the hypothesis of a causality from satisfied material aspirations to cooperation: *if* the person is satisfied income wise *then* her propensity to cooperate will increase, and not viceversa[20].

As argued by Simon (1959) "the notion of satiation plays no role in classical economic theory" (pg. 262) and neither does it in the existing theory of social preferences. In contrast, psychological theories emphasize that the motive to act stems from *drives*, and actions terminate when the drive is satisfied. Early in the forties, renowned psychologist Maslow (1943) proposed a theory of human motivation arguing that humans are motivated by unsatisfied hierarchical needs. In his words:

> "the appearance of one need depends on the prior satisfaction of another, more pre-potent need. [...] No need or drive can be treated as if it were isolated or discrete; every drive is related to the state of satisfaction or dissatisfaction of other drive." (Maslow, 1943, pg. 370)

Maslow's theory is consistent with people having different preference ranking depending on which level of the pyramid they are. A person who is mainly seeking to satisfy lower level needs (e.g. food, water, shelter) may be expected to act "more selfishly" than someone who has overcome her basic needs and is seeking higher goals in the pyramid, such as social goals like sense of belonging, good relationships, etc.

It is argued here that people have *individual material aspirations* that can correspond to the lower steps of Maslow's pyramid and *social aspirations* which can be understood as the upper levels of the pyramid. Each person defines her own aspirations in both dimensions, and H.1 simply states that the marginal rate of substitution between both targets depends on the extent to which the first target is satisfied.

Besides testing H.1, this chapter test three other hypotheses that will provide a better understanding of the situational-dependent social preference argued here.

**H.2 (Social Identity)** $\beta_2 > 0$. Social Identity, at a particular geographic level, positively affects people's propensity to cooperate with fellows of the same geographical

[20] We comment more on this issue of causality in the section of "robustness checks".

level. On average, it is expected that the closer the people feel they are to their neighbour, compatriot or foreign fellow, the higher will be the number of tokens contributed to the local, national or global account, respectively.

As is emphasized by many scholars both at the theoretical (Brewer, 1981; Akerlof and Kranton, 2000) and experimental (Brewer and Kramer, 1986; Schopler and Insko, 1992) level, social identification increase cooperation by reducing actors' tendency to draw distinctions between their own and others' welfare.

It is also known from the sociological and anthropological literature (Appadurai, 2004; Sewell et al., 1957) that individual aspirations are socially determined. As stated by Ray (2006), a person draws her aspirations from the "lives, achievements, or ideals of those who belong to her cognitive world, her zone of similar, attainable individuals, [...], her aspirations neighbourhood."

If it is the case that social identity affects both cooperation and aspirations, and it is also proved that aspirations affect cooperation, then it is essential to include both variables in the same regression to actually learn about each separate effect, if any. Such a procedure has not been followed by the existing literature, since the potential effect of individual aspirations on cooperation has been completely neglected so far.

**H.3 (Income Comparison)** $\beta_3 > 0$. Players' perceptions of their own income relative to the income of the average recipient of the public good, at a local, national or global level, affects propensity to contribute to the local, national or global public good, with a positive sign. That is, the richer people perceive they are relative to their neighbour, compatriot or foreign fellow, the higher will be the number of tokens contributed to the local, national or global account, respectively.

Similar concerns of the effect of distributive justice principles on cooperation have been studied in the Social Psychology literature. Lamm and Schwinger (1980), for example, show empirical evidence that people contribute unequally in favor of the needier person. In the economic literature, a similar version of this hypothesis has been put forward by Fehr and Schmidt (1999), who suggested that some people may be inequity averse and may be willing to sacrifice their own payoff in order to minimize disparities between own and others payoffs. However, H.3 is stating something different. In this experiment, people come to the laboratory already endowed with (a) a particular real income level and (b) a particular perception of their income relative to the income of the average neighbour, compatriot or foreigner. Thus, if H.3 is confirmed, it implies that people's experimental decisions are taken with the motivation of contributing to minimize real income disparities among players, and not only disparities in the payoffs originated in the game. This creates a sharp distinction between what H.3 tests and what is usually tested in the literature of inequality aversion. H.3. claims that a "inequality averse" rich person will cooperate with a poor person, but may not cooperate with a "rich" person. According to Fehr and Schmidt (1999), however, this "inequality averse" person will cooperate in the same way in both experiments, because people's perceptions about their income relative to the others does not play a role in their model. Thus, in the light of their model, the same person will be thought to be pure altruistic in the first experiment, and pure inequality averse in the second.

**H.4 (Interaction)** $\beta_4 < 0$. Players' perception of their relative income w.t. the income of the average recipient of the public good moderates the effect of Income Aspirations Gap on contributions, with a negative sign. In other words, the (negative) effect of income aspirations gap on cooperation is lower (less negative) if the person perceives that her income is higher than the income of the average recipient of the public good.

## 6.3 Results

In order to test H.1, two ordered logistic regressions with *Income Aspirations Gap* as a regressor and total contributions to the "collective" accounts[21] as the dependent variables are performed. The results of these two regressions are shown in Table IV below.

[21] Local and National in Decision N and Local and World in Decision W.

Table IV - Total Cooperation and Aspirations (Decisions N and W)

| Dependent Variable | Total Contributions (Decision N) | Total Contributions (Decision W) |
|---|---|---|
| Income Aspirations' Gap | -0.737 | -1.147*** |
| | (-0.507) | (-0.424) |
| Number of Cases | 197 | 197 |

*Notes: Ordered logistic regression. Each column is a separate regression equation*

*Robust standard errors in parentheses, clustered by Location and income decile*

** p<0.10, ** p<0.05, *** p<0.01*

Despite the sign of income aspirations gap is negative in both regressions as conjectured in H.1, it is only statistically significant in Decision W. This is surprising since both decisions N and W are identical in all senses except for the nationality of the people who comprise the groups that eventually fund a hypothetical public good (i.e. all Argentines in Decision N and 66% of foreigners in Decision W), but we will come back to this point in a moment.

Note also that the negative coefficient of *Income aspirations Gap* in Decision W is considerably high. For a one unit increase in income aspirations gap, a 1.14 decrease in the log odds of the total propensity to cooperate in decision W is expected, ceteris paribus. The predicted probabilities suggest for example that the probability of observing a highly unsatisfied participant (i.e. $g_i(a_i, s_i) = 0.8$) free-riding on a foreigner is 11.5%. In contrast, the probability that a fully satisfied participant (i.e. $g_i(a_i, s_i) = 0$) free-rides on a foreign fellow is 5%. An average participant contributes all her endowments to a public good with probability 28.2% if $g_i(a_i, s_i) = 0$ and 13.6% if $g_i(a_i, s_i) = 0.8$.

This analysis uses as dependent variable the sum of individual contributions in each decision, and as such, it does not provide much information about the surprising feature observed in Table IV. In the following regressions, people's contributions to each collective account (local, national or world) in each decision (N and W) are separately used as dependent variables. Table V (see Appendix) summarizes the results obtained. Aspiration's gaps were found to explain significantly cooperation at the "world" level and also at the "national" level, although it does not affect cooperation with neighbours at "local" level. There may be different interpretations for this, but we will come back to this concern at the end of this part of the section. For the moment, I want to focus attention on the two cases where the coefficients are significantly different from zero, and study the robutness of these results.

The first question that arises is whether it is the **gap** that matters, or is it simply income or/and absolute aspirations levels which affect cooperation. After all, income aspirations gap depends **non-linearly** on both variables. This question can be easily addressed by including both variables, income and aspirations, as regressors. Table V in the appendix confirms that income aspirations gap remains highly significant at the "world" and "national" levels, even after controlling for these two variables and other demographic variables such us location dummies and gender, age and education. Furthermore, absolute aspirations levels and absolute income levels are found to be statistically insignificant. As a consequence, our result suggests that it **is the gap** instead of the absolute income or aspirations per se **which matters** for cooperation with foreigners and compatriots.

The quantitative effect of this result is considerably high. Figure I and Figure II in the appendix illustrate the predicted probabilities of being a free-rider or a fully cooperator with foreigners conditional on *Income Aspirations Gap* and holding fixed absolute income, aspirations and control demographic variables. From Figure II we observe that the probability of contributing nothing to the "world" account is 9.11% if income aspirations gap is $g = 0$, and it is 55.5% if income aspirations gap is very wide, i.e. $g = 0.8$. Figure I shows the probability of contributing all the endowments to the "world" account conditional on income aspirations gap. This probability is 12.8% if income aspirations gap is $g = 0$, and it is 1.2% if $g = 0.8$.

Marginal effects are also computed. For example, the marginal probability of free-riding on foreigners as a function of income aspirations gap:

$$\text{Marginal Effect}: \left.\frac{\partial \Pr(g^{W*} = 0 \,|\, \mathbf{x})}{\partial gap}\right|_{gap=mean(gap)} = 0.38$$

This marginal effect of 0.38 represents the slope of the curve that associates income aspirations gap with the probability of being a 100% free-rider (Figure II) evaluated at the mean of the income aspirations gap (0.16), holding all the other variables constant at their mean level. If the income aspirations gap of an average person increases by 1 percentual point, the probability that she/he will free-ride on foreigners increases by 0.38 percentual points.

A second question is whether the variable of interest, income aspirations gap, is measuring something different than life satisfaction. After all, it could be expected that both variables are highly positive correlated, and we are attributing the effect on cooperation to income aspirations gap wrongly. For that reason, life satisfaction is included to the regressions as another control variable. Table VI shows that our results remain robust, even controlling for life satisfaction.

Finally, it would be interesting to explore why the effect of income aspirations gap on cooperation is conditional on geographical identity. Why does not income aspirations gap affect cooperation with neighbours? We do not have an answer for this question. One conjecture is that people may feel more committed to their neighbours than with their compatriots or foreigners, so they are unconditional cooperators with neighbours to some extent, regardless of how unsatisfied/satisfied their material needs are. However, we don't have data to test that conjecture.

All this discussion can be summarized in the following result:

**Result 1:** *Income aspirations gaps significantly reduce people's propensity to cooperate with foreigners and compatriots, ceteris paribus. However, it does not affect people's propensity to cooperate with neighbours.*

Hypothesis 2 attempts to explore the link between social identity and cooperation. At the same time, including the variable Social Identity (at all geographical levels) as regressors will help to reduce an eventual omitted variable bias in the previous result, since, as it was argued in the preceding section, social identity may mediate the effect of aspirations on cooperation. Table VI (appendix), however, shows that Result 1 is robust to the inclusion of social identity as another regressor. In addition, we observe that H.2 is confirmed at "local" and "world" level, but not at "national" level. The quantitative effect is considerable. The probability that a person who doesn't include the foreigners into her cognitive "world" (i.e. Social identity index at "world" level would be equal to 0 in this case) fully cooperates with them is 14.4%. However, if the person includes the foreigners into her aspirations neighbourhood, then the probability that she fully cooperates is 36.3%[22].

The fact that social identification with compatriots does not affect people's propensity to cooperate with them is at least puzzling. As stated in the previous section, the existing literature argues that social identification increases cooperation by reducing actors' tendency to draw distinctions between their own and the welfare of others. So, it is not clear why this argument would not be valid at any geographical level.

One challenge to this traditional explanation of how identity affects cooperation comes from Yamagishi and Kiyonari (2000). They argue that identity increases cooperation with ingroup members by influencing actors' expectations about fellow ingroup members' behavior. That is, social identity alone is not sufficient to produce in-group favoritism. Group boundaries do not affect cooperation if actors have a more direct basis for forming expectations that others will reciprocate. In light of this argument, it

[22] The effect of Social Identity on cooperation has been already reported in other works (Brewer, 1996, see)

may be the case that Argentineans, regardless of how closely identified they felt with their compatriots, expected less favourable treatment from their compatriots in return. The role of expectations is confirmed when expectations about others' contributions (at all levels) were added to the regression. As can be observed from Table VI (appendix), there is a clear strong relationship between the player's level of cooperation and the expectations of the level of the average partner's cooperativeness. However, social identity remain highly significant at "world" and "local" level, suggesting that Yamagishi and Kiyonari (2000) hypothesis is only confirmed at "national" level. At the other levels, the hypothesis that social identity alone is not sufficient to produce in-group favoritism cannot be rejected.

The discussion above can be summarized as follows:

**Result 2**: *Social Identity at "world" and "local" level significantly increases the propensity to cooperate with foreigners and neighbours, respectively. However, social identity at "national" level does not affects people's propensity to cooperate with compatriots.*

In order to test hypothesis 3, players' own perceptions of their income relative to the average income of their neighbours, compatriots and foreigners are added to the regressions. The conjecture was that people would cooperate more with those groups who they feel are poorer than them. However, I don't find enough evidence to reject the null hypothesis that my conjecture is false. As can be seen in Table VI, the hypothesis that the coefficients of the Relative Income variable are significant from zero cannot be rejected at standards levels of significance. In other words:

**Result 3**: *Players' perception of their relative income with respect to the income of the average recipient of the public good (at a geographical level) does not affect the propensity to cooperate.*

Finally, we test for the existence of an interaction between income aspirations gap and people's perceptions of their income relative to the income of the people who they are playing. After including the interaction term in the four regressions, this interaction effect was found to be different, depending on the identity of the recipients of the public good (Table VI, appendix). The coefficient measuring the interaction is negative and highly significant, as it was conjectured in H.4, only when the recipients of the public good are neighbours. It is positive and significant when the recipients of the public good are foreigners, and it is not significant when they are compatriots. These observations immediately highlight the complexity of human cooperation. The way inequality aversion moderates the effect of income aspirations gap on cooperation is conditional on the geographical identity of the recipients of the public good. On the one hand, if it is a neighbour, then inequality aversion will reduce the negative effect that aspirations have on cooperation. On the other hand, if it is a foreigner, inequality aversion will increase the negative effect that aspirations have on cooperation. In short,

**Result 4**: *Players' perception of their relative income w.t. the income of the average recipient of the public good reduces the effect of Income Aspirations Gap on contributions with their neighbours, and increases the effect of Income aspirations Gap on contributions with foreigners, although it does not affect the effect of income aspiration gaps on cooperation with compatriots.*

Incidentally, it's worth noting that in order to increase the robustness of these four results, socio-demographic controls (e.g. age, gender, education and location dummies) have been added to the four regressions reported here.

To conclude, the main lesson from the results obtained here is that geographical identity of the recipients of a public good affects both, the level of cooperation and the motivations to contribute.

## 6.4 Robustness checks

As in any empirical study, the validity of the results of this chapter are conditional on assumptions made. In this section, those assumptions considered to be the strongest are discussed. The most important threats to internal validity are (a) errors-in-variables bias, (b) simultaneous causality bias and (c) omitted variable bias from a variable that is correlated with the regressors but is unobserved. This latter source of threat has been extensively considered in the preceding section. Thus, this section will focus on exploring the eventual influence of threats (a) and (b) on the results. In addition, model specification and concerns related to the (small) size of the sample used for estimations are also discussed.

### 6.4.1 Errors-in-variables bias

One of the most important assumptions made in any estimation is that the variables are not subject to measurement errors. Arguably, one of the main variables of this study, income aspirations gaps, may suffer from this error. Unfortunately, due to the lack of empirical work on *income aspirations gaps,* there is no other existing measure to contrast with. The closest one is an aspirations/attainment measure used by Easterlin (2003), who took information from a representative survey[23] of the American population that included the following two questions:

> "1. We often hear people talk about what they want out of life. Here are a number of different things. [The respondent is handed a card with a list of 24 items.] When you think of the good life – the life you'd like to have, which of the things on this list, if any, are part of that good life as far as you personally are concerned? 2. "Now would you go down that list and call off all the things you now have?"

[23] These surveys are used in Schor (1998)

The idea behind these two questions is similar to the idea used to measure income aspirations gap in this chapter. The first question gathers information about desires for certain goods, which can be summarized here as desires for a certain income level. The second question tries to capture the place in which the respondents stand in relation to these desires, i.e. to what extent they are fulfilled. This question is analogous to the question on current income made here. So, the only existing measure of "aspirations gaps" is not far from the one used here.

Notwithstanding the comment made above, the measure used here is subject to another concern. The income aspirations people are asked to comment on are not absolute values, but *relative* to the income distribution in Argentina. People are asked to state in which position of the income distribution of their country they would be satisfied. Thus, the variable income aspiration is specifically measuring relative income aspirations. This approach has its positive and negative sides. On the positive side, it is perfectly in line with the "relative income hypothesis" first formulated by Duesenberry (1949). According to this, people essentially care about their relative position rather than their income in absolute terms (see Zizzo and Oswald, 2001). People set their standards relative to the standards of others. On the negative side, such a measure imposes an eventual artificial upper bound on people's aspirations. A person who is already in the highest decile of the income distribution is not allowed to aspire more than this. In order to reduce this measurement error, all the regressions conducted in this study where replicated dropping from the entire sample those subjects who belonged to the highest decile. Table VII in the appendix shows that the results remain qualitatively the same and, moreover, the coefficient of income aspirations gap even become more significant. Likewise, some variables that were not significant before (e.g. income, absolute aspirations and life satisfaction at the national level), now became significant.

### 6.4.2 Simultaneous causality bias

The results of this chapter claim causality instead of simply correlation or association. Therefore, it must be assumed that a reverse direction of the causality from propensity to cooperate to income aspirations gap or social identity is not plausible. How strong is this assumption?

As far as we understand, there is neither existing theory nor empirical evidence or even intuition that supports the claim that because people are more prone to cooperate, they are more satisfied income-wise. However, this does not imply that such a claim can be rejected and, thus, the direction of the causality assumed in this chapter must be taken only as a tentative direction.

Similar analysis can be done with the assumption made that social identity causes in-group cooperation. Although several papers have experimentally proved that social identification causes cooperation (Chen and Li, 2009), the reverse causality has not been disproved yet. It might be the case that the sole fact of cooperating with some groups increases the in-goup identification. An exploration of such concerns, though, is beyond the scope of the present chapter.

### 6.4.3 Model specification and sample size

One of the assumptions underlying ordinal logistic regressions is that the relationship between each pair of outcome groups is the same. In other words, ordinal logistic regression assumes that the coefficients that describe the relationship between, say, the lowest versus all higher categories of the response variable are the same as those that describe the relationship between the next lowest category and all higher categories, etc. This is called the proportional odds assumption or the parallel regression assumption. Because the relationship between all pairs of groups is the same, there is only one set of coefficients (only one model). If this was not the case, we would need different models to describe the relationship between each pair of outcome groups. I test the proportional odds assumption, and I use a Log Likelihood ratio test[24]. How does this likelihood ratio test work? Let there be $m$ outcomes of the dependent variable. The command "omodel" first estimates the ordered Logit model. Then it obtains an approximation of log-likelihood for the ologit model an approximation to the log-likelihood for a completely general model (in which there are different effects at each cut-point for every variable). The difference between the two approximate log-likelihoods is the likelihood-ratio test reported by omodel. The approximation to the log-likelihood value for the general model will usually be less reliable than the first approximation but both are underestimates of the respective true log-likelihoods. Hence in taking the difference between the two approximations a reasonable approximation is assumed to be obtained. The likelihood-ratio test is performed on degrees of freedom that are the difference between the number of fitted parameters in the standard and general models[25]. A significant $p$-value is evidence to reject the null hypothesis that the coefficients are equal across categories. In the first two models of Table VI we found not enough evidence to reject the assumption of proportional odds (model column 1: $chi^2(136) = 147.65$ and $Prob > chi^2 = 0.2334$; model column 2: $chi^2(153) = 163.48$ and $Prob > chi^2 = 0.2664$). In the last two models (col. 3 and 4) the test rejects the null hypothesis at low significance levels, therefore only for these last two models, there is sufficient evidence that the proportional odds assumption does not hold. However, this is not a major concern since the main results of this chapter come from model 1 and 2 in which the assumption of proportional odds ratio is not rejected.

Before finishing this section, it is important to comment on the size of the sample used in this chapter. As an estimation method, ordered logit uses Maximum Likelihood Estimation instead of Least Square Estimation. The maximum likelihood (ML) estimators become minimum variance unbiased estimators as the sample size tends to infinite

[24] stata command: omodel (Wolfe, 1997)

[25] $(p)(m-2)$, where $p$ is the number of regressors specified.

or it is large enough (i.e. the properties hold asymptotically)[26]. Therefore, the ML estimation may be biased and inferential hypothesis tests are uncertain in small sample analyses. The question that arises is what constitutes an adequate sample size. Is a sample of 200 observations big enough? If not, what can be said about the validity of the results of this chapter? Hart and Clark (1999) explore the behavior of ML estimates in probit models across differing sample sizes and with varying numbers of independent variables in Monte Carlo simulations. Their experiments show that (a) the risk of making Type I errors (i.e. rejecting a hypothesis that should have been accepted) does not change appreciably as sample size descends and (b) the risk of making Type II errors (i.e. accepting a hypothesis that should have been rejected) increases dramatically in smaller samples and as the number of regressors increases. These results, in particular the first result, strengthen the validity of the results shown in this chapter. In the analysis made here, the statistical significance of a coefficient is tested and the null hypothesis is that the coefficient are equal to zero. If the risk of "rejecting a null hypothesis when it is actually true" does not change considerably as sample size decreases, then we would expect that those coefficients that are significantly different from zero in this chapter will remain being different from zero if we increased the sample size. It is certainly true, however, that we do not know if some variables that are not statistically significant with the sample size used here, would become statistical significant if largest sample size is used.

## 6.5 Conclusion

Theories of aspirations have been applied to a broad array of issues across the social sciences, including Sociology, Psychology, Anthropology and Decision Theory. Since its

[26] An estimator is unbiased if we take a very large number of random samples with replacement from a population, then the average value of the parameter estimates will be theoretically exactly equal to the population value. An estimator has minimum variance, if it has the smallest variance, and thus the narrowest confidence interval, of all estimators of that type.

inclusion in Economics in the early fifties by Simon (1955, 1959), no systematic work has been carried out to learn about the economic implications of a model concerned with aspirations.

Likewise, human cooperation in a context of private provision of a public good has always been a challenge to social scientists.

This chapter presented the first attempt to explore the link between individual aspirations and cooperation. It combined experimental data with a post-experimental survey to measure the effect of life income aspirations on individual cooperation at different geographical levels of public goods.

It has been shown that the narrower an individual's income aspirations gap, the greater is the propensity to cooperate providing both global and national public goods. However, this effect was not significant in the provision of a local public good. It is the gap that matters, instead of the absolute income or aspirations. The quantitative effect is considerably high. Participants with the widest aspirations gap were 46 percent more likely to free-ride on foreigners than those who were fully satisfied with their income. We found a parallel result suggesting that the effect of social identity on cooperation depends on the geographical level of the public good provided. The more participants feel identified with foreigners or neighbours, the higher is their contribution levels to providing a global and local public good. However, in-group identification with compatriots does not affect participants' contributions to provide a national public good, ceteris paribus.

We have made three contributions to the Economics literature. First, the results found have important implications for the existing knowledge of private provision of public goods, initiated with Bergstrom et al. (1986) influential paper. We suggest that considering the negative effect of income aspirations gap on voluntary contributions to a public good may change their results. In particular, Bergstrom et al. (1986) states that any changes in the wealth distribution that leaves unchanged the aggregate wealth of current contributors will either increase or leave unchanged the equilibrium supply of public good. However, we claim that this result may not hold in the light of our results, since changes in the wealth distribution may have a non-linear effect on people's cooperation through its change in income aspirations gap.

Second, we showed evidence that the motivations to contribute and the level of contributions depend depend on the geographical type of the public good in question. This has practical implications for regional public policies and raises some challenges for the design of policies aiming at increasing individual contribution levels. For example, our chapter suggests that a policy that promotes in-group identification will increase cooperation at a local level but not at a national level. However, a policy aimed at reducing individuals' income aspirations gap will be more effective at increasing both national and global cooperation than local cooperation. In this latter case, one should be cautious, since reducing individual's income aspirations can, on the one hand, increase social welfare by increasing cooperation, but on the other hand, it can reduce social welfare through the effect on growth.

Third, the evidence presented here highlights a previously "unobserved" source of the individual heterogeneity identified in most public good experiments. Taking the example given in the introduction, if an econometrician does not control for Susan's unsatisfied aspirations, then her behaviour in both scenarios will be inferred as random or inconsistent, when in fact it is the case of an omitted variable bias.

The next chapter concludes the thesis and highlights some directions for future research.

# Chapter 7

# Conclusion and Future Research

Grounded on empirical evidence from Psychology, this thesis has relaxed the assumption of perfect reflection made in standard Economics and developed a novel theoretical framework. The theory proposed here is general and falsifiable, and a solution derived from it is ensured to exist under weak assumptions. The main positive and normative features of the framework have been analysed in an individual decision-making context and in an n-person strategic setting. We unified seemingly disconnected models and proposed new applications on individual commitment, sympathy and empowerment.

One of the main dilemmas of Behavioural Economics is whether revealed preferences can still be used for welfare ranking. The framework introduced here provided a suitable locus to investigate this concern and to gain a better understanding of the limits of the revealed preferences approach. Through this investigation, the degree of personal "autonomy" emerged as a natural normative criterion consistent with a vast range of Behavioural Economics models. Even though the concept of autonomy has been widely studied in Philosophy and Psychology, to my knowledge, this is the first time that the concept is brought up to a formal decision-making and game-theoretical framework.

Consistent with autonomy being the normative criterion to adopt, the thesis has argued in favour of different types of interventions that aim to enhance or sustain individual autonomy. Examples of such interventions are cognitive and behavioural therapies that teach how to manage stress, anxiety, temptation and self-confidence or correct maladaptive beliefs that can lead to systematic self-defeating behaviour. Interventions with this type of aims are not only restricted to psychotherapies, though. Social programs aiming at increasing participants' self-confidence or aspirations can also be autonomy supportive. In general, these interventions have the explicit objective of empowering people, i.e. to help them to increase the sense of control over their lives. Policies such as fostering empowerment or emotional intelligence have been largely advocated in other social disciplines but somehow disregarded in the literature of Economics.

This thesis has also proposed a model of aspirations failure where initial disadvantage results in low aspirations and low achievement. The features that characterize a "role model" were studied and the importance of role models for aspirations formation was investigated. The effect of the degree of connectedness in the society on the aspirations formation process was also investigated.

Finally, this thesis reported the first investigation on the empirical link between aspirations and cooperation.

I expect that the analysis presented here represents a substantial original contribution to knowledge as required for the standards of a Ph.D. degree. Inevitably, this thesis leaves some theoretical and empirical questions unanswered. It is my hope, however, that this piece of work is sufficient to suggest that the line of enquiry advocated here is a promising one, and that it will stimulate new undertakings.

I identify several directions for fruitful future research. The main challenge, in my view, concerns the identification and measurement of autonomy. In the near future, I plan to investigate if observed choices alone can be enough to infer whether a person is making decisions autonomously or not. My conjecture is that observed behaviour could provide information about the way the person makes choices. For example, if we observe that a smoker pays for costly treatments to quit smoking but fails to quit, we could reject the hypothesis that her choice to smoke is fully autonomous. This issue has to be further investigated, but in principle, one could look at the choices made by a person to extract further information valid for welfare analysis, which goes beyond the "revealed preference" paradigm. That is, instead of extracting "revealed preferences", we could extract "revealed decision processes" that are informative for welfare purposes[1]. Another interesting idea along this line of thought is to investigate whether it could be advisable to combine choice and non-choice-based approaches to construct a valid welfare measure. This combination may depend on the complexity of the decision in question, on the primitive preferences of the decision-maker and on her extrinsic circumstances.

In relation to how autonomy should be measured, the plan is to develop a measurement theory of autonomy along the lines of Duclos et al.'s (2004) measurement of polarization. The idea is also to review the existing subjective measures already developed by psychologists like Deci and Ryan (2002) and to assess the extent to which their measure corresponds to the concept of autonomy derived from the framework presented in this thesis.

On the applications front, it would be interesting to link more closely our model with the literature of clinical Psychology. My plan is to develop novel applications grounded on evidence-based successful therapies that aim to increase the autonomy of the individual. Likewise, it is in the agenda to apply our framework to introduce the concept of empathy into game theory. The capacity to empathize, that is, to simulate others' internal decision process, should help players to predict and understand other's actions and intentions, with important economic implications. My guess is that the lack of empathy creates potentially avoidable social conflicts. Finally, grounded on the evidence reported in Chapter 6, it would be interesting to develop a model of private provision of a public good at different regional levels that considers the effect of income aspirations gaps, social identity and relative needs. This would help us to better understand the

[1] The approach we plan to pursue is along the lines of Manzini and Mariotti's (2009) approach. The idea is to fully characterize BDPs and test the axioms in the lab against the axioms of SDPs.

effect of these variables on optimal provision of a public good.

This thesis definitively opens new routes to explore in Experimental Economics and Neuroeconomics. In collaboration with Prof. Ghosal and Dr. Mani, I am currently conducting experimental research addressing the factors that affect the aspirations formation process. We are also designing policy experiments with randomized control groups to evaluate the impact of different policies on the autonomy, aspirations and self-confidence of children from disadvantageous backgrounds.

# Appendix A

# Proofs

## A.1 Results from Chapter 2

**Proof. Theorem 1**

This proof is an adaptation of Ghosal (2006) for BDPs.

Recall that the preferences of the decision-maker is denoted by $\succeq$ a binary relation ranking pairs of decision states in $(A \times P) \times (A \times P)$. As the focus is on incomplete preferences, in this section, instead of working with $\succeq$, we find convenient to specify two other preference relations, $\succ$ and $\sim$. The expression $(a, a') \in \succ_p$ is written as $a \succ_p a'$ and is to be read as "$a$ is preferred to $a'$ by the decision-maker when the utility parameter is $p$". Define the sets $\succ_p (a) = \{a' \in A : a' \succ_p a\}$ (the upper section of $\succ_p$), $\succ_p^{-1} (a) = \{a' \in A : a \succ_p a'\}$ (the lower section of $\succ_p$). It is assumed that for each $p \in P$, (i) $\succ_p$ is acyclic i.e. there is no finite set $\left\{a^1, ..., a^T\right\}$ such that $a^{t-1} \succ_p a^t$, $t = 2, ..., T$, and $a^T \succ_p a^1$, and (ii) $\succ_p^{-1} (a)$ is open relative to $A$ i.e. $\succ_p$ has an open lower section[1]. Write $a' \notin \succ_p (a)$ as $a \not\succ_p a'$.

Define a map $\Psi : P \to A$, where $\Psi(p) = \{a' \in A : \succ_p (a') = \emptyset\}$: for each $p \in P$, $\Psi(p)$ is the set of maximal elements of the preference relation $\succ_p$.

Consider the following extension of $\succ$: for each $p$, define $\succeq_p$ on $A$ by

$$a \succeq_p a' \Leftrightarrow a' \not\succ_p a.$$

[1] The continuity assumption, that $\succ_p$ has an open lower section, is weaker than the continuity assumption made by Debreu (1959) (who requires that preferences have both open upper and lower sections), which in turn is weaker than the assumption by Shafer and Sonnenschein (1975) (who assume that preferences have open graphs). Note that assuming $\succ_p$ has an open lower section is consistent with $\succ_p$ being a lexicographic preference ordering over $A$.

As for each $p$, $\succ_p$ is acyclic and therefore irreflexive, it follows that $\succeq_p$ is complete. Let $\hat{\succ}_p$denote the strict preference relation corresponding to $\succeq_p$ i.e. $a\hat{\succ}_p a'$ if and only if $a \succeq_p a'$ but $a' \not\succeq_p a$. For each $p \in P$ and $a, a' \in A$, $a \succ_p a'$ if and only if $a\hat{\succ}_p a'$.
Define a map $\hat{\Psi} : P \rightarrow A$, where $\hat{\Psi}(p) = \{a' \in A : \hat{\succ}_p (a') = \emptyset\}$: for each $p \in P$, $\hat{\Psi}(p)$ is the set of maximal elements of the preference relation $\hat{\succ}_p$.
The following lemma establishes that for each $p$, $\succ_p$ and $\hat{\succ}_p$ are equivalent and have the same set of maximal elements:
**Lemma 1:** For each $p \in P$ and $a, a' \in A$, $a \succ_p a'$ if and only if $a\hat{\succ}_p a'$ and therefore, $\Psi(p) = \hat{\Psi}(p)$.
**Proof:** Fix $p \in P$. Consider a pair $a, a' \in A$ such that $a \succ_p a'$. Then, $a' \not\succ_p a$ and therefore, $a \succeq_p a'$ and as $\succ_p$ is acyclic, $a' \not\succeq_p a$. It follows that $a\hat{\succ}_p a'$. Next, consider a pair $a, a' \in A$ such that $a\hat{\succ}_p a'$. Then, $a \succeq_p a'$ and $a' \not\succeq_p a$. Therefore, $a' \not\succ_p a$ and $a \succ_p a'$. ■
With this result in place the decision problem with incomplete but acyclic preferences is rephrased as a decision problem with complete and acyclic but not necessarily transitive preferences.
Consider the following assumptions:
(A1) $A$ is a compact lattice with the vector ordering[2];
(A2) For each $p$, and $a, a'$, (i) if $a \succeq_p \inf(a, a')$, then $\sup(a, a') \succeq_p a'$ (ii) if $a \succeq_p \sup(a, a')$ then $\inf(a, a') \succeq_p a'$ (quasi-supermodularity);
(A3) For each $a \geq a'$ and $p \geq p'$, (i) if $a \succeq_{p'} a'$, then $a \succeq_p a'$ and (ii) if $a' \succeq_p a$ then $a' \succeq_{p'} a$ (single-crossing property);
(A4) For each $p$ and $a \geq a'$, (i) if $\succ_p (a') = \emptyset$ and $a \succeq_p a'$, then $\succ_p (a) = \emptyset$, and (ii) $\succ_p (a) = \emptyset$ and $a' \succeq_p a$, then $\succ_p (a') = \emptyset$,(monotone closure).
(A5) $P$ is a compact lattice and $\pi$ is a non-decreasing function.
Assumptions (A2)-(A3) are quasi-supermodularity and single-crossing property defined by Milgrom and Shannon (1994). Assumption (A4) is new. It requires that for each $p$, in any mutually unranked pair of vector ordered actions, either both actions are maximal elements of $\succ_p$or neither action is. The role played by assumption (A4) in obtaining the monotone comparative statics with incomplete preferences is clarified by the following example.

[2] A lattice is a partially ordered subset of $\Re^k$ with the vector ordering (the usual component wise ordering: $x \geq y$ if and only if $x_i \geq y_i$ for each $i = 1, .., K$, and $x > y$ if and only if both $x \geq y$ and $x \neq y$, and $x \gg y$ if and only if $x_i > y_i$ for each $i = 1, .., K$). A lattice that is compact (in the usual topology) is a compact lattice.

**Example**: $P$ is single valued and $A$ is the four point lattice in $\Re^2$

$$\{(e, e), (f, e), (e, f), (f, f)\}$$

where $f > e$. Suppose that $(f, f) \succ (e, e)$ but no other pair is ranked. Then, $\Psi$ consists of $\{(f, e), (e, f), (f, f)\}$ clearly not a lattice. Note that in this case, preferences satisfy acyclicity and quasi-supermodularity (and trivially, single-crossing property). However, preferences do not satisfy monotone closure: $(f, e) \geq (e, e)$, with $\succ ((f, e)) = \emptyset$ and $(e, e) \succeq (f, e)$, but $\succ ((e, e)) \neq \emptyset$.

The preceding example demonstrates that with intransitive preferences, quasi-supermodularity on its own, is not sufficient to ensure that the set of maximal elements of $\succ$ is a sublattice of $A$ even when $\succ$ is acyclic. The example also demonstrates that $\succ$ can be acyclic without necessarily satisfying monotone closure and therefore, the two are distinct conditions on preferences.

The following result shows that assumptions (A1)-(A4), taken together, are sufficient to ensure monotone comparative statics with incomplete preferences and together with (A5) ensure the non-emptiness of $E$.

**Theorem.** Under assumptions (A1)-(A4), each $p \in P$, $\Psi(p)$ is non-empty and a compact sublattice of $A$ where both the maximal and minimal elements, denoted by $\bar{a}(p)$ and $\underline{a}(p)$ respectively, are increasing functions on $P$. Moreover, $E \neq \varnothing$.

**Proof.** By assumption, for each $p$, $\succ_p$ is acyclic, $\succ_p^{-1}(a)$ are open relative to $A$ and $A$ is compact. By Bergstrom (1975), it follows that $\Psi(p)$ is non-empty. As Bergstrom (1975) doesn't contain an explicit proof that $\Psi(p)$ is compact, a proof of this claim follows next. To this end, note that the complement of the set $\Psi(p)$ in $A$ is the set $\Psi^c(p) = \{a' \in A : \succ_p (a') \neq \emptyset\}$. If $\Psi^c(p) = \emptyset$, then $\Psi(p) = A$ is necessarily compact. So suppose $\Psi^c(p) \neq \emptyset$. For each $a' \in \Psi^c(p)$, there is $a'' \in A$ such that $a'' \succ_p a'$. By assumption, $\succ_p^{-1}(a'')$ is open relative to $A$. By definition of $\Psi(p)$, $\succ_p^{-1}(a'') \subset \Psi^c(p)$. Therefore, $\succ_p^{-1}(a'')$ is a non-empty neighborhood of $a' \in \Psi^c(p)$ and it is clear that $\Psi^c(p)$ is open and therefore, $\Psi(p)$ is closed. As $A$ is compact, $\Psi(p)$ is also compact. Next, I show that for $p \geq p'$ if $a \in \Psi(p)$ and $a' \in \Psi(p')$, then $\sup(a, a') \in \Psi(p)$ and $\inf(a, a') \in \Psi(p')$. Note that as $a' \in \Psi(p')$, $a' \succeq_{p'} \inf(a, a')$. By quasi-supermodularity, $\sup(a, a') \succeq_{p'} a$. By single-crossing, $\sup(a, a') \succeq_p a$. As $a \in \Psi(p)$, $\succ_p (a) = \emptyset$ and therefore, by monotone closure, as $\sup(a, a') \succeq_p a$, $\succ_p (\sup(a, a')) = \emptyset$ and $\sup(a, a') \in \Psi(p)$. Next, note that as $a \in \Psi(p)$, $a \succeq_p \sup(a, a')$. By single-crossing, $a \succeq_{p'} \sup(a, a')$ and by quasi-supermodularity, $\inf(a, a') \succeq_{p'} a'$. As $a' \in \Psi(p')$, $\succ_{p'} (a') = \emptyset$, and therefore, by monotone closure, as $\inf(a, a') \succeq_{p'} a'$, $\succ_{p'} (\inf(a, a')) = \emptyset$ and $\inf(a, a') \in \Psi(p')$. Therefore, (i) $\Psi(p)$ is ordered, (ii) $\Psi(p)$ is a compact sublattice of $A$ and has a maximal and minimal element (in the usual component wise vector ordering) denoted by $\bar{a}(p)$ and $\underline{a}(p)$, and (iii) both $\bar{a}(p)$ and $\underline{a}(p)$ are increasing functions from $P$ to $A$. Now define a map $\Psi : A \times P \to A \times P$, $\Psi(a,p) = (\Psi_1(p), \Psi_2(a))$ as follows: for each $(a,p)$, $\Psi_1(p) = \{a' \in A : \succ_p (a') = \phi\}$ and $\Psi_2(a) = \pi(a)$. By (A5) it follows that for each $a$, $\pi(a)$ has a maximal and minimal element (in the usual component wise vector ordering) denoted by $\bar{\pi}(a)$ and $\underline{\pi}(a)$ respectively. Therefore, the map $(\bar{a}(p), \bar{\pi}(a))$ is an increasing function from $A \times P$ to itself and as $A \times P$ is a compact (and hence, complete) lattice, by applying Tarski's fix-point theorem, it follows that $(\bar{a}, \bar{p}) = (\bar{a}(\bar{p}), \bar{\pi}(\bar{a}))$ is a fix-point of $\Psi$ and by a symmetric argument, $(\underline{a}(p), \underline{\pi}(a))$ is an increasing function from $A \times P$ to itself and $(\underline{a}, \underline{p}) = (\underline{a}(\underline{p}), \underline{\pi}(\underline{a}))$ is also a fix-point of $\Psi$; moreover, $(\bar{a}, \bar{p})$ and $(\underline{a}, \underline{p})$ are respectively the largest and smallest fix-points of $\Psi$. ■

**Proof. Proposition 2 (Indistinguishability)**

(i) Suppose $(a,p) \in E$. By definition, for all $a' \in A$, $a \succeq_p a'$ for some $p = \pi(a)$. By $(C1)$, for all $a' \in A$, $(a,p) \succeq (a',p')$ for each $p = \pi(a)$ and $p' = \pi(a')$. It follows that $(a,p) \in M$. Next, suppose, by contradiction, $(a,p) \in E \cap M$ but $(C1)$ doesn't hold. As $(a,p) \in E$, for all $a' \in A$, $a \succeq_p a'$ for $p = \pi(a)$. As, by assumption, $(C1)$ doesn't hold there exists $a' \in A$ such that $a \succeq_p a'$ for $p = \pi(a)$ but $(a,p) \prec (a',p')$ for $p = \pi(a)$ and $p' = \pi(a')$. But, then, $(a,p) \notin M$, a contradiction. (ii) Suppose $(a,p) \in M$. As $(a,p) \succeq (a',p')$ for all $(a',p') \in A \times \pi(A)$, by $(C2)$, $(a,p) \succeq (a',p)$ for $p = \pi(a)$. It follows that $(a,p) \in E$. Next, suppose, by contradiction, $(a,p) \in M \cap E$ but $(C2)$ doesn't hold. As $(a,p) \in M$, $(a,p) \succeq (a',p')$ for all $(a',p') \in A \times \pi(A)$. As, by assumption, $(C2)$ doesn't hold, there exists $a' \in A$ such that $a' \succ_p a$ for $p = \pi(a)$. But, then, $(a,p) \notin E$, a contradiction. ■

**Proof. Theorem 2 (Genericity)**

Let $v(a) = u(a, \pi(a))$. The outcome $(\hat{a}, \hat{p}) \in M$ satisfies the first-order condition

$$\partial_a v(\hat{a}) = \partial_a u(\hat{a}, \pi(\hat{a})) + \partial_p u(\hat{a}, \pi(\hat{a}))\partial_a \pi(\hat{a}) = 0 \tag{A.1}$$

while the outcome $(a^*, p^*) \in E$ satisfies the first-order condition

$$\partial_a u(a^*, p^*) = 0, p^* = \pi(a^*). \tag{A.2}$$

For $(a^*, p^*) = (\hat{a}, \hat{p})$, it must be the case that

$$\partial_p u(\hat{a}, \pi(\hat{a}))\partial_a \pi(\hat{a}) = 0. \tag{A.3}$$

It is easily checked that requiring both $(C1)$ and $(C2)$ to hold is equivalent to requiring that equation A.3 also holds. Consider a decision problem with $(a^*, p^*) = (\hat{a}, \hat{p})$. Perturbations of the utility function and the feedback effect do not affect eq. $A.2$ and hence $(a^*, p^*)$ but they do affect eq. $A.3$ and via eq. $A.1$ affect $(\hat{a}, \hat{p})$. Therefore, $(a^*, p^*) \neq (\hat{a}, \hat{p})$ generically. ■

## A.2 Results from Chapter 3

**Proof. Proposition 4**

Suppose for each $a' \in A$ (other than $a$), $a$ is chosen with $a'$ present ($a'$ may be chosen as well). By assumption, for all $a' \in A$, $a \succeq_p a'$ for $p = \pi(a)$. By $(C1)$, for all $a' \in A$, $(a, p) \succeq (a', p')$ for each $p = \pi(a)$ and $p' = \pi(a')$. It follows that any consistent decision state containing $a$ weakly welfare dominates any other decision state containing $a' \neq a$, $a' \in A$. Next, suppose, by contradiction, for each $a' \in A$ (other than $a$), $a$ is chosen with $a'$ present ($a'$ may be chosen as well), but $(C1)$ doesn't hold. By assumption, for all $a' \in A$, $a \succeq_p a'$ for $p = \pi(a)$. As $(C1)$ doesn't hold, there exists $a' \in A$ such that $a \succeq_p a'$ for $p = \pi(a)$ but $(a, p) \prec (a', p')$ for $p = \pi(a)$ and $p' = \pi(a')$, a contradiction. ■

**Proof. Proposition 5**

Since $\pi(a) = P$ for all $a \in A$, any pair $(a, p)$ is a consistent decision state. Thus, for each $p$ if the decision maker solves the decision problem in a behavioral way, for each $A' \subseteq A$, each chosen action will be an element of $c(A', p)$ and therefore, every choice will be a weak individual welfare optimum in $A$. ■

**Proof. Proposition 6**

If $\pi(a) = P$, then $(a, p) \in M$ implies that $(a, p) \succeq (a', p)$ for all $a' \in A$ and therefore, $(a, p) \in E$. That $\hat{c}(A') \subseteq \tilde{c}(A')$ is a direct consequence of $M \subseteq E$. Finally, as $\hat{c}(A') \subseteq \tilde{c}(A')$, $b \notin \tilde{c}(A')$ implies $b \notin \hat{c}(A')$. It follows that if $a \in \hat{W}$ there is no $b \in A$ such that $b\hat{\mathbb{P}}^*a$. Therefore, for every $b \in A$, there exists a non-empty $A' \subseteq A$ with $a, b \in A'$ such that $a \in \hat{c}(A')$. As $\hat{c}(A') \subseteq \tilde{c}(A')$, it also follows that there is no $b \in A$ such that $b\tilde{\mathbb{P}}^*a$.Therefore, $a \in \tilde{W}$ and $\hat{W} \subseteq \tilde{W}$. ■

**Proof. Proposition 7**

As both $A$ and $P$ are finite, and $\succ$ is acyclic, $M$ (which corresponds to the maximal elements of $\succ$) is non-empty. As $M \subseteq E$, so is $E$. ■

**Proof. Proposition 8**

Step 1: We begin by showing that $M \subseteq E$ iff the following condition (condition $C2$) holds: for $(a,p),(a',p') \in A \times P$ such that $(a,p) \succeq (a',p')$, then $(a,p) \succeq (a',p)$ for some $p \in \pi(a)$. Suppose $(a,p) \in M$. As $(a,p) \succeq (a',p')$ for all $(a',p') \in A \times P$ such that $p' \in \pi(a')$, by $(C2)$, $(a,p) \succeq (a',p)$ for some $p \in \pi(a)$. It follows that $(a,p) \in E$. Next, suppose, by contradiction, $M \subseteq E$ but $(C)$ doesn't hold. As $(a,p) \in M$, $(a,p) \succeq (a',p')$ for all $(a',p') \in A \times P$ such that $p' \in \pi(a')$. As, by assumption, $(C2)$ doesn't hold, there exists $a' \in A$ such that $(a',p) \succ (a,p)$ for some $p \in \pi(a)$. But, then, $(a,p) \notin E$, a contradiction.

Step 2: Clearly, if $M \subseteq E$, $\hat{c}(A') \subseteq \tilde{c}(A')$ and by using argument similar to those in Proposition 8, $\hat{W} \subseteq \tilde{W}$. Next, suppose that $\hat{W} \subseteq \tilde{W}$ but $M \subsetneq E$. As $M \subsetneq E$, by step 1 of the argument, it is equivalent to assume that condition $(C2)$ doesn't hold. As $(C2)$ doesn't hold, there exists $a \in A$ such that $(a,p) \succeq (a',p')$ for some $p \in \pi(a)$ and $p' \in \pi(a')$ but $(a,p) \prec (a',p)$ for all $p \in \pi(a)$. It follows that there exists some non-empty $A' \subseteq A$, such that $a \in \hat{c}(A')$ but $a \notin \tilde{c}(A')$ and therefore, $\hat{c}(A') \subsetneq \tilde{c}(A')$. But, then, there exists $b \in A$ (take $b = a'$) and a non-empty subset of $A$, $A'$, such that $a \in \hat{c}(A')$ but $a \notin \tilde{c}(A')$. Therefore, $a \in \hat{W}$ but $a \notin \tilde{W}$, a contradiction. ■

## A.3 Results from Chapter 4

**Proof. Theorem 3**

Lets define an individual psycho-social state as $\sigma_i = (p, a_{-i})$ where $p = (p_1, ..., p_i, ..., p_n) \in P$ and $a_{-i} \in A_{-i}$. The primitives of the model are two maps, $\succ_{i,\sigma_i}: P \times A_{-i} \to A_i \times A_i$ and $\pi_i : A \times P_{-i} \to P_i$. The first map is a preference relation over $A_i$. The expression $(a_i, a_i') \in \succ_{i,\sigma_i}$is written as $a_i \succ_{i,\sigma_i} a_i'$ and is to be read as "$a_i$ is preferred to $a_i'$ when the psycho-social state is $p$ and the actions chosen by other players are $a_{-i}$." Note that in the general model, we assume that preferences of player $i$ are not only determined by her actions and psycho-social states, but also by the psycho-states of the other players. Define the sets $\succ_{i,\sigma_i}(a_i) = \{a_i' \in A_i : a_i' \succ_{i,\sigma_i} a_i\}$ (the upper section of $\succ_{i,\sigma_i}$) and $\succ_{i,\sigma_i}^{-1}(a_i) = \{a_i' \in A_i : a_i \succ_{i,\sigma_i} a_i'\}$ (the lower section of $\succ_{i,\sigma_i}$). We write $a_i' \notin \succ_{i,\sigma_i}(a_i)$ as $a_i \not\succ_{i,\sigma_i} a_i'$ and $a_i' \in \succ_{i,\sigma_i}(a_i)$ as $a_i' \succ_{i,\sigma_i} a_i$. The second map specifies the set of psycho-social states consistent with the actions chosen by each individual and the psycho-social states of the others. Throughout this section, it is assumed that $\pi_{i,a,p_{-i}}$ is non-empty and closed relative to $P$ for each $a$ and $p_{-i}$.

As stated above, a (pure) psycho-social equilibrium is a pair $(a^*, p^*)$ such that for each $i \in N$, $(i)$ given $p^*$ and $a^*_{-i}$, $\succ_{i,\sigma_i^*}(a^*_i) \cap A_i = \varnothing$ where $\sigma^*_i = (p^*, a^*_{-i})$ and $(ii)$ given $a^*$ and $p^*_{-i}$, $p^*_i \in \pi_{i,a^*,p^*_{-i}}$.

Consider the following additional set of assumptions:

*Assumption 1 (AS-1)*: For each $i \in N$, both $A_i$ and $P$ are compact lattices and for each $p_{-i}$ and $a$, $\pi_{i,a,p_{-i}}$ is a compact sublattice of $P$.

*Assumption 2 (AS-2)*: For each $i \in N$, $p$ and $a_{-i}$, $\succ^{-1}_{i,\sigma_i}(a_i)$ is open relative to $A_i$, i.e. $\succ^{-1}_{i,\sigma_i}$ has an open lower section.

*Assumption 3 (AS-3):* For each $i \in N$, $p$ and $a_{-i}$, $\succ^i_{\sigma^i}$ is acyclic, i.e. there is no finite set $\{a^1_i, ..., a^n_i\}$ such that $a^k_i \succ^i_{\sigma^i} a^{k-1}_i$, $k = 2, ..., n$ and $a^1_i \succ^i_{\sigma^i} a^n_i$.

*Assumption 4 (AS-4)*: For each $i \in N$, $p_{-i}$, $a_{-i}$ and $a_i \geq a'_i$, if $p_i \in \pi_{i,a,p_{-i}}$ and $p'_i \in \pi_{i,a'_i,a_{-i},p_{-i}}$, then $\sup(p_i, p'_i) \in \pi_{i,a,p_{-i}}$ and $\inf(p_i, p'_i) \in \pi_{i,a',p_{-i}}$. (psycho-social states of each player are increasing in her actions)

*Assumption 5 (AS-5)*: For each $i \in N$, $p$, $a_{-i}$ and each pair of actions $a_i, a'_i \in A_i$, $(i)$ if $\inf(a_i, a'_i) \not\succ_{i,\sigma_i} a_i$, then $a'_i \not\succ_{i,\sigma_i} \sup(a_i, a'_i)$, and $(ii)$ if $\sup(a_i, a'_i) \not\succ_{i,\sigma_i} a_i$, then $a'_i \not\succ_{i,\sigma_i} \inf(a_i, a'_i)$. (quasi-supermodularity)

*Assumption 6 (AS-6)*: For each $i \in N$, $a_i \geq a'_i$, $a_{-i} \geq a'_{-i}$, $p_i \geq p'_i$ and $p_{-i} \geq p'_{-i}$, $(i)$ if $a'_i \not\succ_{i,a_i,a'_{-i},p} a_i$, then $a \not\succ_{i,\sigma_i} a_i$, $(ii)$ if $a_i \not\succ_{i,\sigma_i} a'_i$, then $a_i \not\succ_{i,a_i,a'_{-i},p} a'_i$, $(iii)$ if $a'_i \not\succ_{i,a,p'_i,p_{-i}} a_i$, then $a'_i \not\succ_{i,\sigma_i} a_i$, $(iv)$ if $a_i \not\succ_{i,\sigma_i} a'_i$, then $a_i \not\succ_{i,a,p'_i,p_{-i}} a'_i$, $(v)$ if $a'_i \not\succ_{i,a,p_i,p'_{-i}} a_i$, then $a'_i \not\succ_{i,\sigma_i} a_i$, $(vi)$ if $a_i \not\succ_{i,\sigma_i} a'_i$, then $a_i \not\succ_{i,a,p_i,p'_{-i}} a'_i$ (single-crossing property in actions and psycho-social states[3])

*Assumption 7 (AS-7)*: For each $i \in N$, $p$, $a_{-i}$ and $a_i \geq a'_i$, $(i)$ if $\succ_{i,\sigma^i}(a'_i) \cap A_i = \varnothing$ and $a'_i \not\succ_{i,\sigma_i} a_i$, then $\succ_{i,\sigma^i}(a_i) \cap A_i = \varnothing$, and $(ii)$ if $\succ_{i,\sigma^i}(a_i) \cap A_i = \varnothing$ and $a'_i \not\succ_{i,\sigma_i} a'_i$, then $\succ_{i,\sigma^i}(a'_i) \cap A_i = \varnothing$ (monotone closure)

*Step 1.* Define a map $\Psi : A \times P \rightarrow A \times P$ as follows:

$\Psi(a, p) = (\Psi^1(a, p), ..., \Psi^i(a, p), ..., \Psi^n(a, p))$, where

$\Psi^i(a, p) = (\Psi^i_1(a_{-i}, p), \Psi^i_2(a, p_{-i}))$ and for each $i$, $a$ and $p$,

$\Psi^i_1(a_{-i}, p) = \left\{a'_i \in A_i : \succ_{i,\sigma^i}(a'_i) \cap A_i = \varnothing\right\}$, and $\Psi^i_2(a, p_{-i}) = \left\{p'_i \in P : p'_i \in \pi_{i,a,p_{-i}}\right\}$.

*Step 2.* We want to show that $\Psi^i_1(a_{-i}, p)$ is a closed and compact sublattice of $A$.

Since for each $i, p$ and $a_{-i}$ $\succ_{i,\sigma^i}$ is acyclic (*AS-3*), $\succ^{-1}_{i,\sigma_i}(a_i)$ is open relative to $A_i$ (*AS-2*) and $A_i$ is compact (*AS-1*), then by Bergstrom (1975), $\Psi^1_i(a_{-i}, p)$ is not empty. Note that the complement of the set $\Psi^i_1(a_{-i}, p)$ in $A_i$ is the set $\Psi^{i,c}_1(a_{-i}, p) = \left\{a'_i \in A_i : \succ_{i,\sigma^i}(a'_i) \cap A_i \neq \varnothing\right\}$. If $\Psi^{i,c}_1(a_{-i}, p) = \varnothing$, then $\Psi^{i,c}_1(a_{-i}, p) = A_i$ is necessarily compact. So suppose $\Psi^{i,c}_1(a_{-i}, p) \neq \varnothing$. For each $a'_i \in \Psi^{i,c}_1(a_{-i}, p)$, there is a $a''_i \in A_i$ such that $a''_i \succ_{i,\sigma^i} a'_i$, i.e. $a'_i \succ^{-1}_{i,\sigma_i}(a''_i)$. By (*AS-2*) $\succ^{-1}_{i,\sigma_i}(a''_i)$ is open relative to $A_i$. By definition of $\Psi^1(a, p)$, $\succ^{-1}_{i,\sigma_i}(a''_i) \subset \Psi^{i,c}_1(a_{-i}, p)$. Therefore, $\succ^{-1}_{i,\sigma_i}(a''_i)$ is a non-empty neighborhood of $a'_i \in \Psi^{i,c}_1(a_{-i}, p)$. Now, consider

[3] Single crossing property is the ordinal analog of increasing differences.

a sequence $\{a_i^k : k \geq 1\}$ such that for each $k \geq 1$, $a_i^k \in \Psi_1^i(a_{-i}, p)$ but $\lim_{k\to\infty} a_i^k = \hat{a}_i \in \Psi_1^{i,c}(a_{-i}, p)$. Now, by assumption, for each $a_i' \in \Psi_1^{i,c}(a_{-i}, p)$, $\succ_{i,\sigma^i}(a_i') \cap A_i$ is open relative to $A_i$ and therefore, there exists a neighborhood of $\hat{a}_i$, $N(\hat{a}_i) \subset \Psi_1^{i,c}(a_{-i}, p)$, a contradiction as there exists a $K > 1$ such that for each $k > K$, $a_i^k \in N(\hat{a}_i)$. It follows that $\hat{a}_i \in \Psi_1^i(a_{-i}, p)$ and therefore, $\Psi_1^i(a_{-i}, p)$ is closed and since $A$ is compact[4] *(AS-1)*, $\Psi_1^i(a_{-i}, p)$ is also compact. Moreover, as $\succ_{i,\sigma^i}$is quasi-supermodular *(AS-5)*, $\Psi_1^i(a_{-i}, p)$ is also ordered and therefore is a compact (and hence, complete) sublattice of $A$. Thus, $\Psi_1^i(a_{-i}, p)$ has a maximal and a minimal element denoted by $\bar{a}_i(a_{-i}, p)$ and $\underline{a}_i(a_{-i}, p)$ respectively.

*Step 3.* Fix $p$. For $a_{-i} \geq a_{-i}'$, let $a_i \in \Psi_1^i(a_{-i}, \cdot)$ and $a_i' \in \Psi_1^i(a_{-i}', \cdot)$. We want to show that $\sup(a_i, a_i') \in \Psi_1^i(a_{-i}, \cdot)$ while $\inf(a_i, a_i') \in \Psi_1^i(a_{-i}', \cdot)$.

First, note that since $a_i' \in \Psi_1^i(a_{-i}', \cdot)$, $\inf(a_i, a_i') \not\succ_{i,a_i,a_{-i}',p} a_i'$. By part $(i)$ of quasi-supermodularity *(AS-5)*, it follows that $a_i \not\succ_{i,a_i,a_{-i}',p} \sup(a_i, a_i')$. By part $(i)$ of single-crossing *(AS-6)*, it follows that $a \not\succ_{i,\sigma^i} \sup(a_i, a_i')$. Since $a_i \in \Psi_1^i(a_{-i}, \cdot)$, $\succ_{i,\sigma^i}(a_i) \cap A_i = \varnothing$ and therefore, by part $(i)$ of monotone closure *(AS-7)* since $a_i \not\succ_{i,\sigma^i} \sup(a_i, a_i')$, $\succ_{i,\sigma^i}(\sup(a_i, a_i')) \cap A_i = \varnothing$. It follows that $\sup(a_i, a_i') \in \Psi_1^i(a_{-i}, \cdot)$.

Next, note that since $a_i \in \Psi_1^i(a_{-i}, \cdot)$, $\sup(a_i, a_i') \not\succ_{i,\sigma^i} a_i$. By part $(ii)$ of single-crossing property, it follows that $\sup(a_i, a_i') \not\succ_{i,a_i,a_{-i}',p} a_i$.By part $(ii)$ of quasi-supermodularity, it follows that $a_i' \not\succ_{i,a_i,a_{-i}',p} \inf(a_i, a_i')$. Since $a_i' \in \Psi_1^i(a_{-i}', \cdot)$, $\succ_{i,a_i,a_{-i}',p}(a_i') \cap A_i = \varnothing$ and therefore, by part $(ii)$ of monotone closure, since $a_i' \not\succ_{i,a_i,a_{-i}',p} \inf(a_i, a_i')$, $\not\succ_{i,a_i,a_{-i}',p} \inf(a_i, a_i') \cap A_i = \varnothing$. It follows that $\inf(a_i, a_i') \in \Psi_1^i(a_{-i}', \cdot)$.

*Step 4.* Fix $a_{-i}$ and $p_{-i}$. For $p_i \geq p_i'$, let $a_i \in \Psi_1^i(a_{-i}, p_i, p_{-i})$ and $a_i' \in \Psi_1^i(a_{-i}, p_i', p_{-i})$. We want to show that $\sup(a_i, a_i') \in \Psi_1^i(p_i, \cdot)$ while $\inf(a_i, a_i') \in \Psi_1^i(p_i', \cdot)$.

First, note that since $a_i' \in \Psi_1^i(p_i', \cdot)$, $\inf(a_i, a_i') \not\succ_{i,a,p_i',p_{-i}} a_i'$. By part $(i)$ of quasi-supermodularity, it follows that $a_i \not\succ_{i,a,p_i',p_{-i}} \sup(a_i, a_i')$. By part $(iii)$ of single-crossing, it follows that $a_i \not\succ_{i,\sigma^i} \sup(a_i, a_i')$. Since $a_i \in \Psi_1^i(p_i, \cdot)$, $\succ_{i,\sigma^i}(a_i) \cap A_i = \varnothing$ and therefore, by part $(i)$ of monotone closure, as $a_i \not\succ_{i,\sigma^i} \sup(a_i, a_i')$, $\succ_{i,\sigma^i}(\sup(a_i, a_i')) \cap A_i = \varnothing$. It follows that $\sup(a_i, a_i') \in \Psi_1^i(p_i, \cdot)$.

Next, note that since $a_i \in \Psi_1^i(p_i, \cdot)$, $\sup(a_i, a_i') \not\succ_{i,\sigma^i} a_i$. By part $(iv)$ of single-crossing property, it follows that $\sup(a_i, a_i') \not\succ_{i,a,p_i',p_{-i}} a_i$. By part $(ii)$ of quasi-supermodularity, it follows that $a_i' \not\succ_{i,a,p_i',p_{-i}} \inf(a_i, a_i')$. Since $a_i' \in \Psi_i^1(p_i', \cdot)$, $\succ_{i,a,p_i',p_{-i}}(a_i') \cap A_i = \varnothing$ and therefore, by part $(ii)$ of monotone closure, as $a_i' \not\succ_{i,a,p_i',p_{-i}} \inf(a_i, a_i')$, $\not\succ_{i,a,p_i',p_{-i}} \inf(a_i, a_i') \cap A_i = \varnothing$. It follows that $\inf(a_i, a_i') \in \Psi_i^1(p_i', \cdot)$.

*Step 5.* Fix $a_{-i}$ and $p_i$. For $p_{-i} \geq p_{-i}'$, let $a_i \in \Psi_1^i(a_{-i}, p_i, p_{-i})$ and $a_i' \in \Psi_1^i(a_{-i}, p_i, p_{-i}')$. We want to show that $\sup(a_i, a_i') \in \Psi_1^i(p_{-i}, \cdot)$ while $\inf(a_i, a_i') \in \Psi_1^i(p_{-i}', \cdot)$.

First, note that since $a_i' \in \Psi_1^i(p_{-i}', \cdot)$, $\inf(a_i, a_i') \not\succ_{i,a,p_i,p_{-i}'} a_i'$. By part $(i)$ of quasi-

[4]If $A_i$ is compact, $A = \prod_{i\in I} A_i$ is also compact.

supermodularity, it follows that $a_i \nsucc_{i,a,p_i,p'_{-i}} \sup(a_i, a'_i)$. By part $(v)$ of single-crossing, it follows that $a_i \nsucc_{i,\sigma^i} \sup(a_i, a'_i)$. Since $a_i \in \Psi^i_1(p_{-i}, \cdot)$, $\succ_{i,\sigma^i} (a_i) \cap A_i = \varnothing$ and therefore, by part $(i)$ of monotone closure, as $a_i \nsucc_{i,\sigma^i} \sup(a_i, a'_i)$ , $\succ_{i,\sigma^i} (\sup(a_i, a'_i)) \cap A_i = \varnothing$. It follows that $(\sup(a_i, a'_i)) \in \Psi^i_1(p_{-i}, \cdot)$.
Next, note that since $a_i \in \Psi^i_1(p_{-i}, \cdot)$, $\sup(a_i, a'_i) \nsucc_{i,\sigma^i} a_i$.By part $(vi)$ of single-crossing property, it follows that $\sup(a_i, a'_i) \nsucc_{i,a,p_i,p'_{-i}} a_i$. By part $(ii)$ of quasi-supermodularity, it follows that $a'_i \nsucc_{i,a,p_i,p'_{-i}} \inf(a_i, a'_i)$. Since $a'_i \in \Psi^1_i(p'_{-i}, \cdot)$, $\succ_{i,a,p_i,p'_{-i}} (a'_i) \cap A_i = \varnothing$ and therefore, by part $(ii)$ of monotone closure, as $a'_i \nsucc_{i,a,p_i,p'_{-i}} \inf(a_i, a'_i)$ , $\nsucc_{i,a,p_i,p'_{-i}}$ $\inf(a_i, a'_i) \cap A_i = \varnothing$. It follows that $\inf(a_i, a'_i) \in \Psi^1_i(p'_{-i}, \cdot)$.
*Step 6.* It follows that both $\bar{a}_i(a_{-i}, p)$ and $\underline{a}_i(a_{-i}, p)$ are increasing in $p$ and in $a_{-i}$. Further, since for each $p_{-i}$ and $a$, $\pi_{i,a,p_{-i}}$ is a compact (and therefore complete) sublattice of $P$ *(AS-1)*, $\Psi^i_2(a, p_{-i})$ has a maximal and a minimal element (in the usual component wise vector ordering): denote these by $\bar{p}_i(a, p_{-i})$ and $\underline{p}_i(a, p_{-i})$ respectively. As $\pi_{i,a,p_{-i}}$ is increasing in $a_i$ *(AS-4)*, both $\bar{p}_i(a, p_{-i})$ and $\underline{p}_i(a, p_{-i})$ are increasing in $a_i$ as well. It follows that the $(\bar{a}_i(a_{-i}, p), \bar{p}_i(a, p_{-i}))$ is an increasing function from $A \times P$ to itself and since $A \times P$ is compact (and hence, complete) lattice, by applying Tarski's fix-point theorem (Tarski, 1955), it follows that $(\bar{a}, \bar{p}) = (\bar{a}_i(\bar{a}_{-i}, \bar{p}), \bar{p}_i(\bar{a}, \bar{p}_{-i}))$ is a fix-point of $\Psi$. By a symmetric argument $(\underline{a}_i(a_{-i}, p), \underline{p}_i(a, p_{-i}))$ is an increasing function from $A \times P$ to itself and therefore, $(\underline{a}, \underline{p}) = (\underline{a}_i(\underline{a}_{-i}, \underline{p}), \underline{p}_i(\underline{a}, \underline{p}_{-i}))$ is also a fix-point of $\Psi$. Moreover,

$$(\bar{a}, \bar{p}) = \sup\{(a, p) \in A \times P : (\bar{a}_i(\bar{a}_{-i}, \bar{p}), \bar{p}_i(\bar{a}, \bar{p}_{-i})) \geq (a, p)\}$$

and

$$(\underline{a}, \underline{p}) = \inf\left\{(a, p) \in A \times P : (\underline{a}_i(\underline{a}_{-i}, \underline{p}), \underline{p}_i(\underline{a}, \underline{p}_{-i})) \leq (a, p)\right\}$$

Therefore, $(\bar{a}, \bar{p})$ and $(\underline{a}, \underline{p})$ are, respectively, the largest and smallest (in the usual component wise vector ordering) fix-points of $\Psi$. ■

**Proof. Theorem 4**

As before, for each $i \in N = \{1, ..., n\}$ there are two sets, a set $A_i$ of pure actions, $A_i \subset \Re^{m_i}$ and a set $P_i$ of psycho-social states, $P_i \subset \Re^{\rho_i}$, where $\Re^{m_i}$ and $\Re^{\rho_i}$ are finite dimensional Euclidian spaces. For each $i \in N$, there is a preference map $\succ_{i,\sigma_i}: P \times A_{-i} \to A_i \times A_i$, where for each $p$ and $a_{-i}$, describes a preference relation over $A_i$ Recall that $P = \prod_{i \in N} P_i$. Let the space of all Borel probability distributions over $P_i$ (respectively $A_i$) be denoted by $\Delta(P_i)$ (respectively $\Delta(A_i)$),for each $i \in N$. As $P_i$ (respectively $A_i$) is separable (in the usual topology) and $\Delta(P_i)$ (respectively $\Delta(A_i)$) is endowed with the topology of weak convergence, $\Delta(P_i)$ (respectively $\Delta(A_i)$) is separable and metrizable by the Lévy-Prokhorov metric[5].

Let $\hat{\succ}_i : \Delta(P_i) \times \Delta(P_{-i}) \times \Delta(A_{-i}) \to \Delta(A_i) \times \Delta(A_i)$, be a map that describes preferences over probability distributions over $P$ and $A_{-i}$ where for each $\mu \in \Delta(P_i) \times \Delta(P_{-i})$ and $s_{-i} \in \Delta(A_{-i})$, $\hat{\succ}_{i,\mu,s_{-i}}$describes a preference relation over $\Delta(A_i)$. For $s_i$ and $s_i'$ $\in \Delta(A_i)$,the expression $(s_i, s_i') \in \hat{\succ}_{i,\mu,s_{-i}}$is written as $s_i \hat{\succ}_{i,\mu,s_{-i}} s_i'$ and is to be read as "$s_i$ is preferred to $s_i'$ by player $i$ when the distribution over the set of psycho-social states and the actions chosen by other players is $\mu_i$ and $s_{-i}$ respectively." For any set $X$, note that the set of Dirac probability measures[6] over $X$ is simply $X$ itself. Let $\hat{\succ}_i^D : P \times A_{-i} \to A_i \times A_i$ denote the restriction of $\hat{\succ}_i$to Dirac probability measures. We assume the preferences of player $i$ over probability distributions are consistent with her preferences over pure actions, i.e. $\hat{\succ}_i^D \equiv \succ_i$. As before, define the sets $\hat{\succ}_{i,\mu,s_{-i}}(s_i) = \left\{ s_i' \in A_i : s_i' \hat{\succ}_{i,\mu,s_{-i}} s_i \right\}$ (the upper section of $\hat{\succ}_{i,\mu,s_{-i}}$) and $\hat{\succ}_{i,\mu,s_{-i}}^{-1}(s_i) = \left\{ s_i' \in A_i : s_i \hat{\succ}_{i,\mu,s_{-i}} s_i' \right\}$ (the lower section of $\hat{\succ}_{i,\mu,s_{-i}}$). We write $s_i' \notin \hat{\succ}_{i,\mu,s_{-i}}(s_i)$ as $s_i \hat{\not\succ}_{i,\mu,s_{-i}} s_i'$ and $s_i' \in \hat{\succ}_{i,\mu,s_{-i}}(s_i)$ as $s_i' \hat{\succ}_{i,\mu,s_{-i}} s_i$.

As before, there is a map $\pi_i : A \times P_{-i} \to P_i$ that specifies the set of psycho-social states consistent with the actions chosen by each individual and the psycho-social states of the others. In this part of the chapter, it is assumed that $\pi_{i,a,p_{-i}}$ is non-empty for each $a \in A$ and $p_{-i} \in P_{-i}$ and $\pi_i$ is a continuous function on $A \times P_{-i}$.

A *(mixed) psycho-social equilibrium* is a pair $(s^*, \mu^*)$ such that for each $i \in N$, $(i)$ given $\mu^*$ and $s_{-i}^*$, $\hat{\succ}_{i,\mu^*,s_{-i}^*}(s_i^*) \cap A_i = \varnothing$ and $(ii)$ for each $i \in N$, and for each $p$ such that $\mu^*(p) >> 0$, there is $a \in A$ such that $s^*(a) >> 0$ and $\mu_i^*(p_i) = \mu_i^*(\pi_i(a^*, p_{-i}^*)) = s^*(a)$.

We make the following assumptions:

*Assumption 1' (AS-1'):* For each $i \in N$, both $A_i$ and $P$ are compact and hence, $\Delta(A_i)$ and $\Delta(P)$ are compact sets.

*Assumption 2' (AS-2')*: For each $i \in N$, $\mu$ and $s_{-i}$, both $\hat{\succ}_{i,\mu,s_{-i}}$and $\hat{\succ}_{i,\mu,s_{-i}}^{-1}$are open relative to $\Delta(A_i)$, i.e. $\hat{\succ}_{i,\mu,s_{-i}}(s_i)$ has both open upper and lower sections.

*Assumption 3' (AS-3')*: For each $i \in N$, $\mu$ and $s_{-i}$, $\hat{\succ}_{i,\mu,s_{-i}}$is acyclic, i.e. there is no finite set $\{s_i^1, ..., s_i^n\}$ such that $s_i^k \hat{\succ}_{i,\mu,s_{-i}} s_i^{k-1}$, $k = 2, ..., n$ and $s_i^1 \hat{\succ}_{i,\mu,s_{-i}} s_i^n$.

*Assumption 4' (AS-4')*: For each $i \in N$, $\mu$, $s_{-i}$, $s_i$ and $s_i'$, if $\hat{\succ}_{i,\mu,s_{-i}}(s_i) \cap A_i \neq \varnothing$ and $\hat{\succ}_{i,\mu,s_{-i}}(s_i') \cap A_i \neq \varnothing$, then $\hat{\succ}_{i,\mu,s_{-i}}(\lambda s_i + (1-\lambda)s_i') \cap A_i \neq \varnothing$ for each $\lambda \in [0,1]$ (convexity)

*Step 1'*. Define a map $\hat{\Psi} : \Delta(A) \times \Delta(P) \to \Delta(A) \times \Delta(P)$,

$\hat{\Psi}(s, \mu) = \left( \hat{\Psi}^1(s, \mu), ..., \hat{\Psi}^i(s, \mu), ..., \hat{\Psi}^n(s, \mu) \right)$, where

$\hat{\Psi}^i(s, \mu) = \left( \hat{\Psi}_1^i(s_{-i}, \mu), \hat{\Psi}_2^i(s, \mu_{-i}) \right)$ and for each $i \in N$, $s$ and $\mu$,

[5]The Lévy-Prokhorov metric is a metric (i.e. a definition of distance) on the collection of probability measures on a given metric space. For more details see Billingsley (1999).

[6]The Dirac measure is a measure $\delta x$ on a set $X$ (with any sigma algebra of subsets of $X$) that gives the singleton set $\{x\}$ the measure 1, for a chosen element $x \in X$.

$$\hat{\Psi}_1^i(s_{-i},\mu) = \left\{s_i' \in \Delta(A_i) : \hat{\succ}_{i,\mu,s_{-i}}(s_i') \cap A_i = \varnothing\right\}$$

and

$$\hat{\Psi}_2^i(s,\mu_{-i}) = \left\{ \begin{array}{c} \mu_i \in \Delta(P_i) : \mu^*(p) >> 0, \text{ iff } \exists\ a \in A \\ \text{s.t. } s^*(a) >> 0 \text{ and } \mu_i^*(p_i) = \mu_i^*(\pi_i(a^*, p_{-i}^*)) = s^*(a). \end{array} \right\}$$

*Step 2'.* Using a similar argument to the one used in Theorem 3, we know that $\hat{\Psi}_1^i(s_{-i},\mu)$ is non-empty and compact.

*Step 3'.* Now we want to show that $\hat{\Psi}_1^i(s_{-i},\mu)$ is upper semi-continuous. As the range of $\hat{\Psi}_1^i(s_{-i},\mu)$ is compact, $\hat{\Psi}_1^i(s_{-i},\mu)$ is upper semi-continuous if $\hat{\Psi}_1^i(s_{-i},\mu)$ has the closed graph property. Consider four convergent sequences $\left\{s_i^\tau, s_{-i}^\tau, \mu_i^\tau, \mu_{-i}^\tau : \tau \geq 1\right\}$ such that $\lim_{\tau\to\infty} s_i^\tau = \hat{s}_i$, $\lim_{\tau\to\infty} s_{-i}^\tau = \hat{s}_{-i}$, $\lim_{\tau\to\infty} \mu_i^\tau = \hat{\mu}_i$, $\lim_{\tau\to\infty} \mu_{-i}^\tau = \hat{\mu}_{-i}$ and for each $\tau \geq 1$, $s_i^\tau \in \hat{\Psi}_1^i(s_{-i}^\tau, \mu^\tau)$, with $\mu^\tau = (\mu_i^\tau, \mu_{-i}^\tau)$ but $\hat{s}_i \notin \hat{\Psi}_1^i(\hat{s}_{-i},\hat{\mu})$, with $\hat{\mu} = (\hat{\mu}_i, \hat{\mu}_{-i})$ i.e. $\hat{s}_i \in \Psi_1^{i,c}(\hat{s}_{-i},\hat{\mu}) = \left\{s_i' \in \Delta(A_i) : \hat{\succ}_{i,\mu,s_{-i}}(s_i') \cap A_i \neq \varnothing\right\}$. Again, by the assumption that $\hat{\succ}_{i,\mu,s_{-i}}(\hat{s}_i)$ has an open lower section (AS-2'), arguments similar to those used in Theorem 3 show that there exists a non-empty neighborhood of $\hat{s}_i$, $N(\hat{s}_i) \subset \Psi_1^{i,c}(\hat{s}_{-i},\hat{\mu})$, a contradiction as there exists a $\bar{\tau} > 1$ such that for each $\tau > \bar{\tau}$, $s_i^\tau \in N(\hat{s}_i)$. Now for each $s_i' \notin \hat{\Psi}_1^i(\hat{s}_{-i},\hat{\mu})$, there is a $s_i'' \in \Delta(A_i)$ such that $s_i' \hat{\succ}_{i,\hat{\mu},\hat{s}_{-i}} s_i''$ i.e. $s_i' \in \hat{\succ}_{i,\hat{\mu},\hat{s}_{-i}}(s_i'')$. By assumption (AS-2'), $\hat{\succ}_{i,\hat{\mu},\hat{s}_{-i}}(s_i'')$ is open relative to $\Delta(A_i)$ and therefore there is a neighborhood $N(s_i') \subset \hat{\succ}_{i,\hat{\mu},\hat{s}_{-i}}(s_i'')$. As $s_i^\tau \in \hat{\Psi}_1^i(s_{-i}^\tau, \mu^\tau)$, there is some $s_i' \in \hat{\Psi}_1^i(\hat{s}_{-i},\hat{\mu})$ and $\breve{\tau} \geq 1$ such that for all $\tau > \breve{\tau}$, $s_i^\tau \in N(s_i')$ and therefore, $s_i'$ is a limit point of the sequence $\{s_i^\tau : \tau \geq 1\}$, a contradiction as all the subsequences of convergent sequence must have the same limit. It follows that $\hat{\Psi}_1^i(s_{-i},\mu)$ has the closed graph property. Moreover, by (AS-4') $\hat{\Psi}_1^i(s_{-i},\mu)$ is also convex.

*Step 4'.* By the continuity of the map $\pi_i(.)$, $\hat{\Psi}_2^i(s,\mu_{-i})$ is also a continuous function.

*Step 5'.* It follows that $\hat{\Psi}$ satisfies all the assumption of Fan-Glicksberg fix-point theorem and therefore has a fix-point $(s^*,\mu^*)$, which by construction, is a random psycho-social equilibrium. ■

**Proof. Proposition 9**

$(\mathbf{C},\mathbf{C})$ is a "reciprocity-fairness" equilibrium iff for $i = 1,2$:

$x_i(\mathbf{C},\mathbf{C}) + \rho_i\left[\tilde{f}_j(\mathbf{C},\mathbf{C})f_i(\mathbf{C},\mathbf{C})\right] \geq x_i(\mathbf{D},\mathbf{C}) + \rho_i\left[\tilde{f}_j(\mathbf{C},\mathbf{C})f_i(\mathbf{D},\mathbf{C})\right] \Leftrightarrow 4X + \rho_i[\frac{1}{2}\frac{1}{2}] \geq 6X + \rho_i[\frac{1}{2}(-\frac{1}{2})]$

$\Leftrightarrow \rho_i \geq 4X$

$(\mathbf{C},\mathbf{C})$ is a "guilt-aversion" equilibrium iff for $i = 1,2$:

$x_i(\mathbf{C},\mathbf{C}) - \gamma_i \max\{0, x_j(\mathbf{C}) - x_j(\mathbf{C})\} \geq x_i(\mathbf{D},\mathbf{C}) - \gamma_i \max\{0, x_j(\mathbf{C}) - x_j(\mathbf{D})\} \Leftrightarrow 4X \geq 6X - 4X\gamma_i$

$\Leftrightarrow \gamma_i \geq \frac{1}{2}$

And it can be similarly shown that $(\mathbf{C},\mathbf{C})$ is a "commitment" equilibrium. ■

**Proof. Proposition 10**

The proof of this proposition is a direct extension of the proof of Theorem 2, Ch.2 ■

**Proof. Proposition 11**

Let's define an appropriate best-response equivalent transformation of $\Gamma$ as one in which:

$$\hat{x}_i(a_i, a_{-i}) = \begin{cases} x_i(a_i, a_{-i}) + B_i, \text{ for all } a_i \in A_i \text{ given } a^*_{-i} \in A_{-i}, \text{ for all } i \in N, \\ x_i(a_i, a_{-i}), \qquad\qquad\qquad\qquad\qquad\qquad \text{otherwise} \end{cases}$$

where $B_i = -\pi_i(a^*_i, a^*_{-i})$ and $a^* = (a^*_i, a^*_{-i})$ is a pure strategy Nash-equilibrium profile of $\Gamma$.

For example, suppose that the original normal form game ($\Gamma$) is:

| | L | R |
|---|---|---|
| U | $\underline{2}$,-$\underline{1}$ | $\underline{2}$,-2 |
| D | 1,$\underline{2}$ | 1,1 |

So, computing the appropriate transformation of $\Gamma$, we have $\hat{\Gamma} := \{A_1, ..., A_n; \hat{v}_1, ..., \hat{v}_n\}$ which is strategically equivalent to $\Gamma$. In this case, the transformed normal form game $\hat{\Gamma}$ is:

| | L | R |
|---|---|---|
| U | $\underline{0}$,$\underline{0}$ | $\underline{2}$,-1 |
| D | -1,$\underline{2}$ | 1,1 |

Note that $\Gamma$ and $\hat{\Gamma}$ are *best-response equivalent* but they are not fully-equivalent since players' preferences change with the transformation. For instance, player 1 prefers (U,L) to (R,D) in $\Gamma$ but prefers (R,D) to (U,L) in $\hat{\Gamma}$.

Continuing with our proof for any normal form game, we shall show that $a^* = (a^*_i, a^*_{-i})$ is a pure strategy Nash-equilibrium of the restricted $\hat{\Gamma}$ for any restriction.

A pure strategy Nash-Equilibrium (N.E) of the normal form game $\hat{\Gamma}$, $s^* = (s_i^*, s_{-i}^*)$ is such that:

$\hat{v}_i(s_i^*, s_{-i}^*) \geq \hat{v}_i(s_i', s_{-i}^*), \forall i \in N, \forall s_i' \in S_i$, iff

$$\sum_{a_{-i} \in A_{-i}} \sum_{a_i \in A_i} s_{-i}^*(a_{-i}) s_i^*(a_i) \hat{x}_i(a_i, a_{-i}) \geq \sum_{a_{-i} \in A_{-i}} \sum_{a_i \in A_i} s_{-i}^*(a_{-i}) s_i'(a_i) \hat{x}_i(a_i, a_{-i}), \forall i \in N$$

By definition of a pure strategy N.E, $\exists\ a^*\ s.t.\ s_i^*(a_i^*) = 1$ with $s_i^*(a_i') = 0, \forall i \in N, \forall a_i' \neq a_i^*, a_i' \in A_i$. Then, $\hat{x}_i(a_i^*, a_{-i}^*) \geq \sum_{a_i \in A_i} s_i'(a_i) \hat{x}_i(a_i, a_{-i}^*)$, and we know that, given the particular transformation we have defined, $\hat{x}_i(a_i^*, a_{-i}^*) = 0$.

Now let's analyze the equilibrium of $\hat{\Gamma}$ for all possible restrictions.

Case 1: Assume that $\pi_i(s_i, s_{-i}) = \sum_{a_i \in A_i} s_i'(a_i) \hat{x}_i(a_i, a_{-i}^*)$, $\forall i \in N$

Then, $s_i^*(a_i^*) = 1\ \forall i \in N$ is a (weak) equilibrium of the restricted $\hat{\Gamma}$.

Case 2: Assume that $\pi_i(s_i, s_{-i}) = s_i(a_i^*) s_{-i}(a_{-i}^*) \hat{x}_i(a_i^*, a_{-i}^*)$, $\forall i \in N$

Then, $s_i^*(a_i^*) = 1\ \forall i \in N$ is a (strict) equilibrium of the restricted $\hat{\Gamma}$

Case 3: Assume that $\pi_i(s_i, s_{-i}) \neq s_i(a_i^*) s_{-i}(a_{-i}^*) \hat{x}_i(a_i^*, a_{-i}^*)$, $\forall i \in N$

Then, $s_i^*(a_i^*) = 1\ \forall i \in N$ is a (weak or strict) equilibrium of the restricted $\hat{\Gamma}$ ■

**Proof. Proposition 12**

To prove this proposition, it is enough to show that there exists some non-trivial restriction such that a non-trivial non-degenerate mixed strategy equilibrium of the restricted normal form game ($G'$) does not exist. For example, take any arbitrary $a_i' \in A_i$ for each $i \in N$. Then, consider the following restriction:

$$\pi_i(s_i, s_{-i}) = \sum_{a_{-i} \in A_{-i}} \sum_{a_i \in A_i} s_{-i}(a_{-i}) s_i(a_i) x_i(a_i, a_{-i}), \forall a_i \backslash a_i', \forall i \in N$$

We can show that there is not a non-trivial normal form game under $\pi_i(s_i, s_{-i})$ restriction, that could support a non-trivial non-degenerate mixed strategy equilibrium. $s^* = (s_i^*, s_{-i}^*)$ is a non-degenerate mixed strategy equilibrium of $G'$ with $\pi_i(s_i, s_{-i})$ as restriction iff:

$(i)$ $s_{-i}(a_{-i}') x_i(a_i', a_{-i}') = 0$, for $a_i' \in A_i, \forall i \in N$ and

$(ii)$ $s_i(a_i) \in (0, 1), \forall i \in N$ and $\forall a_i \in A_i$

For these two conditions to hold, we need $x_i(a_i', a_{-i}') = 0$. But if this is the case, the restricted game is trivial. ■

**Proof. Proposition 13**

To prove this proposition, it is sufficient to pick a game and show that there is an outcome that is not an equilibrium for all non-trivial consistent restriction. For example, take the following prisoners' dilemma (p.d.) game:

| | | | player 2 | |
|---|---|---|---|---|
| | | | $q$ | $(1-q)$ |
| | | | C | D |
| player 1 | $p$ | C | 1,1 | -1,2 |
| | $(1-p)$ | D | 2,-1 | $\varepsilon$,$\varepsilon$ |

where $-1 < \varepsilon < 0$. We shall show that any outcome with mixed strategies $q = p \in (\frac{1}{2}, 1]$ can not be rationalized as an equilibrium of any non-trivial consistent restricted game derived from p.d. game. Take the following 6 non-trivial consistent restrictions:

1) $\pi_1^1(s_1, s_2) = (-1)(1-q)$ and $\pi_2^1(s_1, s_2) = (-1)(1-p)$

$(p^*, q^*)$ is an equilibrium of the restricted game iff: $q^* = 2q^* + \varepsilon(1-q^*)$ and $p^* = 2p^* + \varepsilon(1-p^*)$.Then $p^* = q^* = \frac{\varepsilon}{\varepsilon-1}$. Since $-1 < \varepsilon < 0$, then $p^* = q^* \in (0, \frac{1}{2})$.

2) $\pi_1^2(s_1, s_2) = q$ and $\pi_2^2(s_1, s_2) = p$

$(p^*, q^*)$ is an equilibrium of the restricted game iff: $(-1)(1-q^*) = 2q^* + \varepsilon(1-q^*)$ and $(-1)(1-p^*) = 2p^* + \varepsilon(1-p^*)$.Then $p^* = q^* = \frac{\varepsilon-1}{\varepsilon+1}$. Since $-1 < \varepsilon < 0$, then $p^* = q^* < 0$.

3) $\pi_1^3(s_1, s_2) = q + (-1)(1-q)$ and $\pi_2^3(s_1, s_2) = p + (-1)(1-p)$

$(p^*, q^*)$ is an equilibrium of the restricted game iff: $0 = 2q^* + \varepsilon(1-q^*)$ and $0 = 2p^* + \varepsilon(1-p^*)$.Then $p^* = q^* = \frac{\varepsilon}{\varepsilon-2}$. Since $-1 < \varepsilon < 0$, then $p^* = q^* \in (0, \frac{1}{3})$.

4) $\pi_1^4(s_1, s_2) = 2q$ and $\pi_2^4(s_1, s_2) = 2p$

$(p^*, q^*)$ is an equilibrium of the restricted game iff: $q + (-1)(1-q^*) = \varepsilon(1-q^*)$ and $p^* + (-1)(1-p^*) = \varepsilon(1-p^*)$.Then $p^* = q^* = \frac{\varepsilon+1}{\varepsilon+2}$. Since $-1 < \varepsilon < 0$, then $p^* = q^* \in (0, \frac{1}{2})$.

5) $\pi_1^5(s_1, s_2) = \varepsilon(1-q)$ and $\pi_2^5(s_1, s_2) = \varepsilon(1-p)$. $\nexists\, q \in [0,1]$ and a $p \in [0,1]$ s.t. each player is indifferent between her actions.

6) $\pi_1^6(s_1, s_2) = 2q + \varepsilon(1-q)$ and $\pi_2^6(s_1, s_2) = 2p + \varepsilon(1-p)$

$(p^*, q^*)$ is an equilibrium of the restricted game iff: $q - 1(1-q) = 0$ and $p - 1(1-p) = 0$.Then $p^* = q^* = 0$.

The rest of the consistent restrictions are linear combinations of the six restrictions above. It is easy to show that such restrictions generated by a linear combination will give a $\hat{q}^* = \hat{p}^*$ for a fixed $-1 < \varepsilon < 0$ which is lower than the maximum $q^* = p^*$(i.e. lower than

$\frac{1}{2}$). Take for example the lineal combination of $\pi_1^1(s_1, s_2)$ and $\pi_1^2(s_1, s_2)$ as $\pi_1^{1,2}(s_1, s_2; z) = z(\pi_1^1(s_1, s_2)) + (1-z)(\pi_1^2(s_1, s_2))$, for $0 \leq z \leq 1$. Then, $z(-1+q) + (1-z)q = -q + \varepsilon q$ iff $q^* = \frac{-z}{\varepsilon - 2}$. Since $0 \leq z \leq 1$ and $-1 < \varepsilon < 0$, then $\hat{q}^* = \hat{p}^* \in [0, \frac{1}{2}] < q^* = p^*$ for a fixed value of $-1 < \varepsilon < 0$. Analogously, we can show that this is the case for all possible linear combinations of restrictions for all $i$. ■

**Proof. Proposition 14**

If $a \in A$ is a pure strategy Nash equilibrium of the original game, then given $a_{-i}$, $a_i \in \arg\max u_i(a_i, a_{-i})$, for all $i \in N$ and $a_i \in A_i$ and $a_{-i} \in A_{-i}$. Moreover, it must be that given $p_i = (a_i, a_{-i})$ and $a_{-i}$, $a_i \in \arg\max v_i(a_i, a_{-i}, p_i)$ for all $i \in N$ and $a_i \in A_i$ and $a_{-i} \in A_{-i}$. This is true, since in a loss aversion psycho-social equilibrium, $u_i(p_i) = u_i(a_i, a_{-i})$, and by the characteristics of the reference-dependent utility function $v_i$, this implies that $v_i(a_i, a_{-i}, a_i, a_{-i}) = u_i(a_i, a_{-i})$. Therefore, $a \in A$ is also a strategy profile of a loss aversion psycho-social equilibrium $(a, a)$ under approach 1. To show that the proposition holds also for non-myopic loss aversion equilibrium, if there is a gain from an unilateral deviation in pure strategies for a loss averse non-myopic player, $v_i(a_i', a_{-i}, a_i', a_{-i}) > v_i(a_i, a_{-i}, a_i, a_{-i})$, which implies $v_i(a_i', a_{-i}, a_i', a_{-i}) > u_i(a_i, a_{-i})$ and $u_i(a_i', a_{-i}) > u_i(a_i, a_{-i})$, a contradiction, since $a \in A$ is a pure strategy Nash equilibrium. ■

## A.4 Results from Chapter 5

**Proof. Proposition 15**

Let $\hat{\Delta}(p, \theta_0)$ be the value of $\Delta$ that solves the equation $pb(\Delta(p, \theta_0) + \theta_0) - \overline{c}(\theta_0) = 0$. Evidently, as the LHS in eq. 5.2 is increasing in $\Delta$, eq. 5.2 holds iff $\Delta \geq \hat{\Delta}(p, \theta_0)$. Moreover, (a) as expected payoffs are increasing in $p$, a higher value of $p$ implies a lower value of $\hat{\Delta}(p, \theta_0)$ is required to ensure that the equation continues to be satisfied, and (b) as $b(.)$ is (weakly) increasing in $\theta_0$ and $\overline{c}(\theta_0)$ is decreasing in $\theta_0$, a higher value of $\theta_0$ implies a lower value of $\hat{\Delta}(p, \theta_0)$ is required to ensure that the equation continues to be satisfied. Therefore, the probability that the decision-maker will choose $\overline{a}$ is $1 - G(\hat{\Delta}(p, \theta_0))$. As $\hat{\Delta}(p, \theta_0)$ is decreasing in both $p$ and $\theta_0$ and $G(.)$ is a cumulative distribution function (and therefore, an increasing function), it follows that $G(\hat{\Delta}(p, \theta_0))$ is decreasing in both $p, \theta_0$ as well. Therefore, (i) given $\theta_0$, higher values of $p$ are associated with a higher likelihood that the individual will choose $\overline{a}$;(ii) given $p$, higher values of $\theta_0$ are associated with a higher likelihood that the individual will choose $\overline{a}$. ■

**Proof. Proposition 16**

Fix $\Delta, \theta_0$. Let $\hat{p}(\Delta, \theta_0)$ be the value of $p$ that solves the equation $p(\Delta, \theta_0)\, b(\Delta + \theta_0) - \overline{c}(\theta_0) = 0$. Evidently, as the LHS in eq. 5.2 is increasing in $p$, eq. 5.2 holds iff $p_0 \geq \hat{p}(\Delta, \theta_0)$. Moreover, (a) as $b(.)$ is increasing in $\Delta$, a higher value of $\Delta$ implies a lower value of $\hat{p}(\Delta, \theta_0)$ is required to ensure that the equation continues to be satisfied, and (b) as $b(.)$ is (weakly) increasing in $\theta_0$ and $\overline{c}(\theta_0)$ is decreasing in $\theta_0$, a higher value of $\theta_0$ implies a lower value of $\hat{p}(\Delta, \theta_0)$ is required to ensure that the equation continues to be satisfied. Therefore, the probability that the decision-maker will choose $\overline{a}$ is $1 - F(\hat{p}(\Delta, \theta_0))$. As $\hat{p}(\Delta, \theta_0)$ is decreasing in both $\Delta$ and $\theta_0$, and $F(.)$ is a cumulative distribution function (and therefore, an increasing function), it follows that $F(\hat{p}(\Delta, \theta_0))$ is decreasing in both $\Delta$ and $\theta_0$. Therefore, (i) given $\theta_0$, higher values of $\Delta$ are associated with a higher likelihood that the individual will choose $\overline{a}$;(ii) given $\Delta$, higher values of $\theta_0$ are associated with a higher likelihood that the individual will choose $\overline{a}$. ■

**Proof. Proposition 17**

The decision-maker will anticipate that she will choose $\overline{a}$ iff

$$p_0 b(\Delta' + \theta_0) - \overline{c}(\theta_0) \geq 0$$

$\Leftrightarrow$

$$p_0 \geq \hat{p}(\Delta', \theta_0).$$

Therefore, the decision-maker will choose $\Delta = 0$ if $p_0 < \hat{p}(\Delta', \theta_0)$. ■

**Proof. Proposition 19**

Suppose the decision maker chooses a cognitive neighborhood $[\underline{\theta}, \underline{\theta} + d]$. Then, the probability that the decision-maker will observe a signal $\left(\theta^j, \theta_1^j = \Delta^{j\prime} + \theta_0^j\right)$ for each $\theta_0^j \in [\underline{\theta}, \underline{\theta} + d]$ is $1 - F(\hat{p}\left(\Delta^{j\prime}, \theta_0^j\right)$ and therefore, her posterior conditional on $\left(\theta_0^j, \theta_1^j = \Delta^{j\prime} + \theta_0^j\right)$ is

$$p_1\left(\theta_0^j\right) = \left(1 - F(\hat{p}\left(\Delta^{j\prime}, \theta_0^j\right))\right) \frac{s\left(\theta_0^j, \theta_0\right) + p_0}{1 + s\left(\theta_0^j, \theta_0\right)} + F(\hat{p}\left(\Delta^{j\prime}, \theta_0^j\right)) p_0.$$

Clearly, the decision-maker will choose $\overline{a}$ iff $p_1\left(\theta_0^j\right) \geq \hat{p}(\Delta', \theta_0)$ or equivalently iff

$$\frac{\left(1 - F(\hat{p}\left(\Delta^{j\prime}, \theta_0^j\right))\right)}{F(\hat{p}\left(\Delta^{j\prime}, \theta_0^j\right))} \left(\frac{s\left(\theta_0^j, \theta_0\right) + p_0}{1 + s\left(\theta_0^j, \theta_0\right)}\right) \geq \frac{\hat{p}(\Delta', \theta_0)}{F(\hat{p}\left(\Delta^{j\prime}, \theta_0^j\right))} - p_0$$

By the continuity of the similarity function $s\left(\theta_0^j, \theta_0\right)$, $\frac{s\left(\theta_0^j, \theta_0\right) + p_0}{1 + s\left(\theta_0^j, \theta_0\right)}$ is decreasing and continuous in $\theta_0^j$ while as $\frac{(1-F(x))}{F(x)}$ is decreasing and continuous in $x$, $\frac{\left(1-F\left(\hat{p}\left(\Delta^{j\prime},\theta_0^j\right)\right)\right)}{F\left(\hat{p}\left(\Delta^{j\prime},\theta_0^j\right)\right)}$ is decreasing and continuous in $\theta_0^j$. Therefore, there exists $\bar{d}>0$ such that for all $\theta_0^j \in [\underline{\theta}, \underline{\theta}+d]$, $p_1\left(\theta_0^j\right) \geq \hat{p}\left(\Delta', \theta_0\right)$ while for all $\theta_0^j > \underline{\theta}+\bar{d}$, $p_1\left(\theta_0^j\right) < \hat{p}\left(\Delta', \theta_0\right)$. It follows that there exists $\varepsilon>0$ such that for all $\theta_0^j \in \left(\overline{\theta}-\varepsilon, \overline{\theta}\right]$,

$$\frac{\left(1-F\left(\hat{p}\left(\Delta^{j\prime},\theta_0^j\right)\right)\right)}{F\left(\hat{p}\left(\Delta^{j\prime},\theta_0^j\right)\right)}\left(\frac{s\left(\theta_0^j,\theta_0\right)+p_0}{1+s\left(\theta_0^j,\theta_0\right)}\right) \leq \frac{\hat{p}\left(\Delta',\theta_0\right)}{F\left(\hat{p}\left(\Delta^{j\prime},\theta_0^j\right)\right)} - p_0$$

Therefore, if $d^* = \left|\overline{\theta}-\underline{\theta}\right|$, with probability $\frac{\left|\overline{\theta}-\varepsilon,\overline{\theta}\right|}{\left|\overline{\theta}-\underline{\theta}\right|}$ the decision-maker doesn't revise her actions. By choosing $d=\left|\overline{\theta}-\varepsilon,\overline{\theta}\right| < \left|\overline{\theta}-\underline{\theta}\right|$, the decision-maker reduces the cognitive cost from $C(\left|\overline{\theta}-\underline{\theta}\right|)$ to $C(\left|\overline{\theta}-\varepsilon,\overline{\theta}\right|)$ without impacting on expected payoffs at Step 2 of the decision-process, a net utility gain. Note that as $\frac{\left(1-F\left(\hat{p}\left(\Delta^{j\prime},\theta_0^j\right)\right)\right)}{F\left(\hat{p}\left(\Delta^{j\prime},\theta_0^j\right)\right)}\left(\frac{s\left(\theta_0^j,\theta_0\right)+p_0}{1+s\left(\theta_0^j,\theta_0\right)}\right)$ is decreasing and continuous in $\theta_0^j$, if

$$p_1\left(\theta_0^j\right) < \hat{p}\left(\Delta',\theta_0\right)$$

for some $\theta_0^j$ then

$$p_1\left(\theta_0^{jj}\right) < \hat{p}\left(\Delta',\theta_0\right)$$

for all $\theta_0^{jj} \geq \theta_0^j$: it follows that the decision-maker will choose a subinterval of $\left[\underline{\theta},\overline{\theta}\right]$. Finally, note that the expected gross payoff gain from choosing the cognitive window $d \leq \bar{d}$ is

$$\int_{\underline{\theta}}^{\underline{\theta}+d} p_1\left(\theta_0^j\right)\left(\frac{b\left(\Delta'+\theta_0\right)}{d}\right) d\theta_0^j - c\left(\theta_0\right)$$

Hence, the optimal choice $d^* \leq \bar{d}$ will solve the inequality

$$\int_{\underline{\theta}}^{\underline{\theta}+d^*} p_1\left(\theta_0^j\right)\left(\frac{b\left(\Delta'+\theta_0\right)}{d^*}\right) d\theta_0^j - c\left(\theta_0\right) \geq C(d^*)$$

with equality if $d^* < \bar{d}$. ■

**Proof. Proposition 20**

Consider an individual $i$ located at $\theta_0^i$ choosing $a=\underline{a}$ with aspiration level $\Delta=0$. Suppose for some $0<\varepsilon$ there exists an $j \neq i$ such that $\left|\theta_0^i-\theta_0^j\right| < \varepsilon$ and $j$ chooses an $a=\overline{a}$ and aspiration level $\Delta=\Delta'$. Then, given a continuous monotone similarity function $s^i(.,.)$ there exists $0<\varepsilon^i$ such that if $\left|\theta_0^i-\theta_0^j\right|<\varepsilon$, $0<\varepsilon<\varepsilon^i$, $p_0^i < \hat{p}\left(\Delta',\theta_0^i\right) \leq$

$\frac{s(\theta_0^j,\theta_0^i)+p_0^i}{1+s(\theta_0^j,\theta_0^i)}$ i.e. the $j$ belongs to the cognitive neighborhood chosen by $i$. Let $\bar{\varepsilon} = \min_i \varepsilon^i > 0$. Then, as long as the distribution of individuals in a society is $\varepsilon-$dense, $0 < \varepsilon < \bar{\varepsilon}$, the society is strongly connected. Now suppose there is one individual with $k$ with $p_0^k \geq \hat{p}\left(\Delta', \theta_0^k\right)$. This individual $k$ will choose $a = \overline{a}$ and aspiration level $\Delta = \Delta'$. Moreover, as the society is strongly connected, there is a path linking $k$ to every other individual $j$, $j \neq k$ i.e. there is a finite chain of individuals $i_0, i_1, ..., i_{\hat{n}}$ with $i_0 = k$ and $i_{\hat{n}} = j$ such that $i_n$ belongs to the cognitive neighborhood of $i_{n+1}$, $n = 0, ..., \hat{n} - 1$. Therefore, each $i_n$, $n = 0, ..., \hat{n}$, chooses $a = \overline{a}$ and aspiration level $\Delta = \Delta'$. More generally, the probability that all individuals in it choose $\overline{a}$ and have aspirations $\Delta'$ is $\max_{k \in N} 1 - F(\hat{p}\left(\Delta', \theta_0^k\right))$. ■

**Proof. Proposition 21**

Note that the adaptive dynamics will always converge to either a type I or a type II outcome. Further, note that the basin of attraction for a type I outcome is $[0, \hat{\Delta}(p, \theta_0))$ while the basin of attraction for a type II outcome is $(\hat{\Delta}(p, \theta_0), K]$. Therefore, the probability that the dynamics will converge to a type I outcome is $G(\hat{\Delta}(p, \theta_0))\,(1-p)$ while the probability that the dynamics will converge to a type II outcome is $\left(1 - G(\hat{\Delta}(p, \theta_0))\right) p$. As $\hat{\Delta}(p, \theta_0)$ is decreasing in $\theta$, it follows that there exists a $\tilde{\theta}$ such that whenever
(a) $\theta_0 < \tilde{\theta}$, $G(\hat{\Delta}(p, \theta_0))\,(1-p) > \left(1 - G(\hat{\Delta}(p, \theta_0))\right)$ and a type I outcome will have a higher probability of emerging while
(b) $\theta_0 > \tilde{\theta}$, $G(\hat{\Delta}(p, \theta_0))\,(1-p) < \left(1 - G(\hat{\Delta}(p, \theta_0))\right)$ and a type II outcome will have a higher probability of emerging. ■

# Appendix B

# Figures and Tables

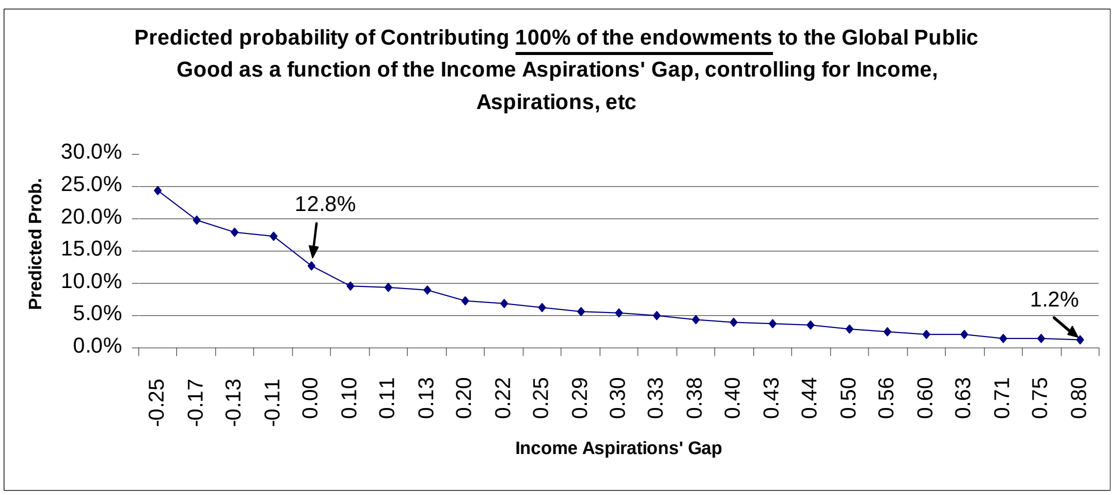


Figure B.1: Predicted Probability of Contributing all endowments to the GPG

**TABLE III - Descriptive Statistics of Independent Variables**

| Range | | mean | sd | max | min |
|---|---|---|---|---|---|
| [1-4] | Income Aspiration Gap | 0.16 | 0.22 | 0.80 | -1.00 |
| [1-10] | Absolute Aspirations Level | 8.56 | 1.72 | 10.00 | 1.00 |
| [1-10] | Income | 7.30 | 2.43 | 10.00 | 1.00 |
| [0-1] | Social Identity (at World Level) | 0.51 | 0.27 | 1.00 | 0.00 |
| [0-1] | Social Identity (at National Level) | 0.79 | 0.19 | 1.00 | 0.11 |
| [0-1] | Social Identity (at Local Level) | 0.76 | 0.24 | 1.00 | 0.00 |
| [1-5] | Relative income (at World Level) | 2.18 | 1.00 | 1.00 | 5.00 |
| [1-5] | Relative income (at National Level) | 3.08 | 1.03 | 1.00 | 5.00 |
| [1-5] | Relative income (at Local Level) | 2.92 | 0.88 | 1.00 | 5.00 |
| [0-10] | Expectations about others' contribution (at World Level) | 5.17 | 2.83 | 0.00 | 10.00 |
| [0-10] | Expectations about others' contribution (at National Level) | 5.26 | 2.56 | 0.00 | 10.00 |
| [0-10] | Expectations about others' contribution (at Local Level) | 4.77 | 2.45 | 0.00 | 10.00 |
| [1-4] | Life Satisfaction | 3.15 | 0.73 | 4.00 | 1.00 |
| [1-6] | Education | 2.61 | 1.06 | 5.00 | 1.00 |
| | Gender | 1.57 | 0.50 | 2.00 | 1.00 |
| | Age | 39.42 | 11.98 | 75.00 | 18.00 |

**TABLE V - Result I**

| **Dependent Variable: Number of tokens contributed** | **Decision W Global PG** | **Decision N National PG** | **Decision N Local PG** | **Decision W Local PG** |
|---|---|---|---|---|
| **Income Aspirations' Gap (IAG)** | **-3.139*****  | **-2.811*** | -0.48 | 1.067 |
| | (1.217) | (1.436) | (2.254) | (1.793) |
| Income | -0.32 | -0.345 | -0.072 | 0.148 |
| | (0.197) | (0.225) | (0.368) | (0.264) |
| Absolute Aspirations Level | 0.221 | 0.211 | -0.026 | -0.237 |
| | (0.189) | (0.228) | (0.36) | (0.238) |
| Location dummies included | Yes | Yes | Yes | Yes |
| Demographic variables included (education, age, gender) | Yes | Yes | Yes | Yes |
| Number of Cases | 194 | 194 | 194 | 194 |

*Notes: Ordered logistic regression. Each column is a separate regression equation*
*Robust standard errors in parentheses, clustered by Location and income decile*
* $p<0.10$, ** $p<0.05$, *** $p<0.01$; *PG = Public Good*

**TABLE VI - Complete Regressions**

| | **Dependent Variable: Number of tokens contributed** | **Decision W Global PG** | **Decision N National PG** | **Decision N Local PG** | **Decision W Local PG** |
|---|---|---|---|---|---|
| **Result 1** | **Income Aspirations' Gap** | **-3.949*** | **-9.446*** | -0.929 | 2.265 |
| | | (2.259) | (5.063) | (3.047) | (1.889) |
| | Income | -0.013 | -0.993 | -0.598 | -0.304 |
| | | (0.308) | (0.662) | (0.386) | (0.286) |
| | Absolute Aspirations' Level | -0.046 | 0.718 | 0.591 | 0.294 |
| | | (0.244) | (0.588) | (0.364) | (0.251) |
| **Result 2** | **Social Identity (World Level)** | **1.251**** | 0.603 | **-1.305**** | **-1.944***** |
| | | (0.521) | (0.93) | (0.599) | (0.66) |
| | Social Identity (National level) | -0.711 | 0.266 | -0.634 | -0.196 |
| | | (0.819) | (0.91) | (0.522) | (0.385) |
| | **Social Identity (Local level)** | -0.164 | -0.221 | **2.019***** | **2.413***** |
| | | (0.849) | (0.799) | (0.648) | (0.547) |
| **Result 3** | Relative income (World level) | -0.037 | | | |
| | | (0.295) | | | |
| | Relative income (National level) | | 0.171 | | |
| | | | (0.234) | | |
| | Relative income (Local level) | | | 0.235 | 0.204 |
| | | | | (0.225) | (0.191) |
| **Result 4** | **Relative income (World level) x IAG** | **1.492*** | | | |
| | | (0.855) | | | |
| | Relative income (National level) x IAG | | 0.407 | | |
| | | | (0.913) | | |
| | **Relative income (Local level) x IAG** | | | **-1.285**** | **-1.589***** |
| | | | | (0.578) | (0.543) |
| **Controls** | Life Satisfaction | -0.205 | -0.173 | 0.096 | 0.068 |
| | | (0.229) | (0.216) | (0.282) | (0.216) |
| | **Expectations on others' contribut (World Level)** | **0.454***** | | | |
| | | (0.066) | | | |
| | **Expectations on others' contribution (National Level)** | | **0.487***** | | |
| | | | (0.087) | | |
| | **Expectations on others' contribution (Local Level)** | | | **0.238**** | **0.397***** |
| | | | | (0.093) | (0.09) |
| | Location dummies included | Yes | Yes | Yes | Yes |
| | Demographic variables included (education, age, gender) | Yes | Yes | Yes | Yes |
| | Number of Cases | 174 | 172 | 179 | 177 |

*Notes: Ordered logistic regression. Each column is a separate regression equation*
*Robust standard errors in parentheses, clustered by Location and income decile*
* $p<0.10$, ** $p<0.05$, *** $p<0.01$

**TABLE VII: Complete Regressions (observations at the 10th Decile dropped)**

| | Dependent Variable: Number of tokens contributed | Decision W Global PG | Decision N National PG | Decision N Local PG | Decision W Local PG |
|---|---|---|---|---|---|
| **Result 1** | **Income Aspirations' Gap** | **-6.079**** | **-13.505***** | -0.944 | 1.703 |
| | | (2.449) | (5.213) | (3.352) | (2.376) |
| | Income | -0.042 | **-1.316**** | -0.647 | -0.208 |
| | | (0.373) | (0.661) | (0.419) | (0.254) |
| | Absolute Aspirations' Level | -0.016 | **0.959*** | 0.638 | 0.241 |
| | | (0.254) | (0.559) | (0.391) | (0.237) |
| **Result 2** | **Social Identity (World Level)** | **1.663***** | 0.879 | -1.058 | **-2.176**** |
| | | (0.603) | (1.013) | (0.735) | (0.848) |
| | Social Identity (National level) | -0.851 | -0.217 | -0.901 | -0.436 |
| | | (0.959) | (0.983) | (0.655) | (0.476) |
| | **Social Identity (Local level)** | -0.291 | 0.294 | **2.081***** | **2.573***** |
| | | (1.08) | (0.93) | (0.672) | (0.677) |
| **Result 3** | Relative income (World level) | -0.411 | | | |
| | | (0.325) | | | |
| | Relative income (National level) | | -0.033 | | |
| | | | (0.283) | | |
| | Relative income (Local level) | | | 0.312 | 0.023 |
| | | | | (0.242) | (0.204) |
| **Result 4** | **Relative income (World level) x IAG** | **2.679***** | | | |
| | | (0.869) | | | |
| | Relative income (at National level) x IAG | | 1.379 | | |
| | | | (0.986) | | |
| | **Relative income (at Local level) x IAG** | | | **-1.544***** | **-1.334**** |
| | | | | (0.596) | (0.619) |
| **Controls** | **Life Satisfaction** | -0.33 | **-0.393*** | 0.16 | 0.189 |
| | | (0.224) | (0.205) | (0.326) | (0.242) |
| | **Expectations on others' contribution (World Level)** | **0.466***** | | | |
| | | (0.086) | | | |
| | **Expectations on others' contribution (National Level)** | | **0.481***** | | |
| | | | (0.108) | | |
| | **Expectations on others' contribution (Local Level)** | | | **0.239**** | **0.484***** |
| | | | | (0.116) | (0.079) |
| | Location dummies included | Yes | Yes | Yes | Yes |
| | Demographic variables included (education, age, gender) | Yes | Yes | Yes | Yes |
| | Number of Cases | 136 | 135 | 140 | 138 |

*Notes: Ordered logistic regression. Each column is a separate regression equation*
*Robust standard errors in parentheses, clustered by Location and income decile*
* $p<0.10$, ** $p<0.05$, *** $p<0.01$

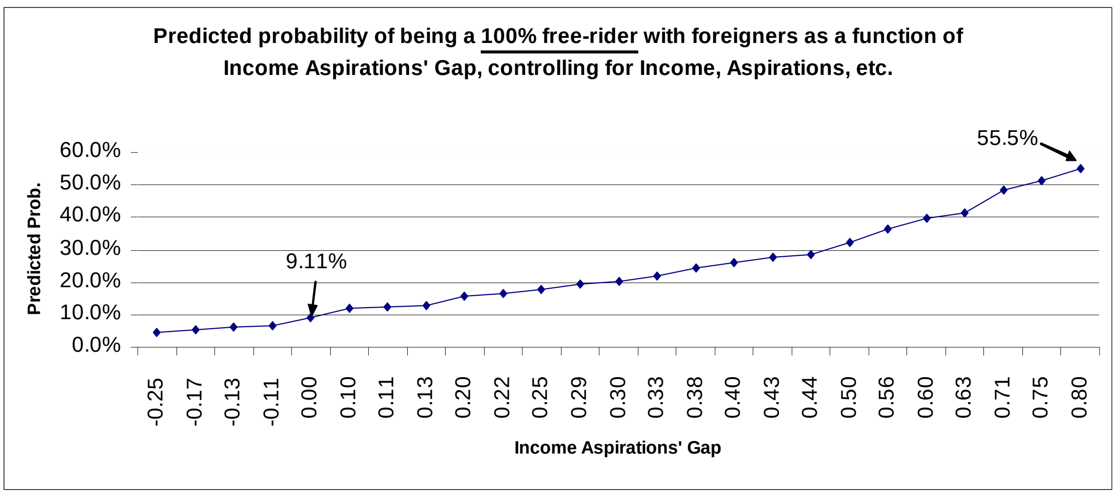


Figure B.2: Predicted Probability of Being free-rider

# Bibliography


Akerlof, G. and W. Dickens (1982). The Economic Consequences of Cognitive Dissonance. *The American Economic Review 72*(3), 307–319.

Akerlof, G. A. and R. E. Kranton (2000). Economics and Identity. *The Quarterly Journal of Economics 115*(3), 715–753.

Appadurai, A. (2004). The Capacity to Aspire. In V. Rao and M. Walton (Eds.), *Culture and Public Action*, pp. 59–84. Washington, DC: The World Bank.

Ariely, D. and K. Wertenbroch (2002). Procrastination, deadlines, and performance: Self-control by precommitment. *Psychological Science 13*(3), 219–224.

Arrow, K. (1959). Rational Choice Functions and Orderings. *Economica 26*(102), 121–127.

Atkinson, A. (2003). Income Inequality in OECD Countries: Data and Explanations. *CESifo Economic Studies 49*(4), 479–513.

Atkinson, A. B. (1998). Social Exclusion, Poverty and Unemployment. In A. B. Atkinson and J. Hills (Eds.), *Exclusion, Employment and Opportunity*. LSE, London, UK: Centre for Analysis of Social Exclusion.

Aumann, R. (1997). Rationality and Bounded Rationality. *Games and Economic Behavior 21*, 2–14.

Azariadis, C. (2004). The Theory of Poverty Traps: What have we Learned. In *Poverty Traps*. NJ, US: Princeton University Press.

Azariadis, C. and J. Stachurski (2005). Poverty traps. In P. Aghion and S. Durlauf (Eds.), *Handbook of Economic Growth*, pp. 295–384. North-Holland.

Bandura, A. (1986). *Social Foundations of Thought and Action: A Social Cognitive Theory*. Prentice Hall.

Bandura, A. (1997). *Self-efficacy: The Exercise of Control*. New York: Worth Publishers.

Bandura, A. (2000). Self-efficacy: The Foundation of Agency. In W. J. Perrig (Ed.), *Control of Human Behavior, Mental Processes, and Consciousness*, pp. 17–33. Erlbaum.

Bandura, A. (2001). Social Cognitive Theory: An Agentic Perspective. *Annual Reviews in Psychology 52*(1), 1–26.

Banerjee, A. and A. Newman (1991). Risk-Bearing and the Theory of Income Distribution. *Review of Economic Studies 58*(2), 211–35.

Banerjee, A. and A. Newman (1993). Occupational Choice and the Process of Development. *Journal of Political Economy 101*(2), 274.

Bardhan, P. (1997). Corruption and Development: A Review of Issues. *Journal of Economic Literature 35*(3), 1320–1346.

Baron, J. (2006). *Thinking and Deciding*. Cambridge University Press.

Baumeister, R. and S. Scher (1988). Self-defeating Behavior Patterns among Normal Individuals: Review and Analysis of Common Self-destructive Tendencies. *Psychological Bulletin 104*(1), 3–22.

Bendor, J., D. Mookherjee, and D. Ray (2001). Aspiration-based Reinforcement Learning in Repeated Interaction Games: An Overview. *International Game Theory Review 3*(2-3), 159–174.

Bergstrom, T. (1975). Maximal Elements of Acyclic Relations on Compact Sets. *Ted C Bergstrom 10*(3), 71.

Bergstrom, T., L. Blume, and H. Varian (1986). On the Private Provision of Public Goods. *Journal of Public Economics 29*(1), 25–49.

Bernheim, B. and A. Rangel (2004). Addiction and Cue-triggered Decision Processes. *American Economic Review 94*(5), 1558–1590.

Bernheim, B. and A. Rangel (2008). Beyond Revealed Preference: Choice Theoretic Foundations for Behavioral Welfare Economics. *NBER Working Paper No 13737*.

Bernheim, B. and A. Rangel (2009). Beyond Revealed Preference: Choice-Theoretic Foundations for Behavioral Welfare Economics. *Quarterly Journal of Economics 124*(1), 51–104.

Bernheim, D. and A. Rangel (2005). Behavioral Public Economics: Welfare and Policy Analysis with Non-standard Decision-makers. In *Economic Institutions and Behavioral Economics*, pp. 73–87.

Bertrand, M., S. Mullainathan, and E. Shafir (2004). A Behavioral-Economics View of Poverty. *American Economic Review 94*(2), 419–423.

Beshears, J., J. Choi, D. Laibson, and B. Madrian (2008). How are Preferences Revealed? *Journal of Public Economics 92*(8-9), 1787–1794.

Biewen, M. (2006). Who are the Chronic Poor? An Econometric Analysis of Chronic Poverty in Germany. *Research on Economic Inequality 13*, 31–62.

Billingsley, P. (1999). *Convergence of Probability Measures*. New York: John Wiley & Sons.

Blackwell, C. and M. McKee (2003). Only for my Own Neighborhood? Preferences and Voluntary Provision of Local and Global Public Goods. *Journal of Economic Behavior and Organization 52*(1), 115–131.

Bracha, A. and D. Brown (2008). Affective Decision Making: a Behavioral Theory of Choice. *Yale ICF Working Paper No. 04-03*.

Brewer, M. (1981). Ethnocentrism and its Role in Interpersonal Trust. *Scientific inquiry and the social sciences 214*, 345–360.

Brewer, M. (1996). When Contact is not Enough: Social Identity and Intergroup Cooperation. *International Journal of Intercultural Relations 20*(3-4), 291–303.

Brewer, M. and R. Kramer (1986). Choice Behavior in Social Dilemmas: Effects of Social Identity, Group Size, and Decision Framing. *Journal of Personality and Social Psychology 50*(3), 543–549.

Buchan, N., G. Grimalda, R. Wilson, M. Brewer, E. Fatas, and M. Foddy (2009). Globalization and Human Cooperation. *Proceedings of the National Academy of Sciences 106*(11), 4138.

Caplin, A. and J. Leahy (2001). Psychological expected utility theory and anticipatory feelings. *The Quarterly Journal of Economics 116*(1), 55–79.

Carrillo, J. D. and T. Mariotti (2000). Strategic ignorance as a self-disciplining device. *The Review of Economic Studies 67*(3), 529–544.

Charness, G. and M. Dufwenberg (2006). Promises and Partnership. *Econometrica 74*(6), 1579–1601.

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

Chen, Y. and S. Li (2009). Group identity and social preferences. *American Economic Review 99*(1), 431–457.

Da Rin, M. and T. Hellmann (2002). Banks as Catalysts for Industrialization. *Journal of Financial Intermediation 11*, 366–397.

Dasgupta, P. and D. Ray (1986). Inequality as a Determinant of Malnutrition and Unemployment: Theory. *The Economic Journal 96*(384), 1011–1034.

Davidson, D. (2001). *Essays on Actions and Events*. Oxford University Press.

Dawes, R. M. and R. H. Thaler (1988). Anomalies: Cooperation. *The Journal of Economic Perspectives 2*(3), 187–197.

Deaton, A. (2001). Relative Deprivation, Inequality, and Mortality. *NBER Working Paper No 8099*.

Debreu, G. (1959). *Theory of Value: an Axiomatic Analysis of Economic Equilibrium*. Yale University Press.

Deci, E. and R. Ryan (2000). The" What" and" Why" of Goal Pursuits: Human Needs and the Self-Determination of Behavior. *Psychological Inquiry 11*(4), 227–268.

Deci, E. and R. Ryan (2002). *Handbook of Self-determination Research*. University of Rochester Press.

Duclos, J., J. Esteban, and D. Ray (2004). Polarization: Concepts, Measurement, Estimation. *Econometrica 72*(6), 1737–1772.

Duclos, J. and P. Grégoire (2002). Absolute and Relative Deprivation and the Measurement of Poverty. *Review of Income and Wealth 48*, 471–492.

Duesenberry, J. (1949). *Income, Saving, and the Theory of Consumer Behavior*. Harvard University Press.

Duflo, E. (2006). Poor but Rational. In A. Banerjee, R. Benabou, and D. Mookherjee (Eds.), *Understanding Poverty*. Oxford, UK: Oxford University Press.

Duncan, G., B. Gustafsson, R. Hauser, G. Schmauss, H. Messinger, R. Muffels, B. Nolan, and J. Ray (1993). Poverty Dynamics in Eight Countries. *Journal of Population Economics 6*(3), 215–234.

Durlauf, S. (2006). Groups, Social Influences and Inequality. In *Poverty Traps*, pp. 141–175. Princeton University Press.

Dworkin, G. (1988). *The Theory and Practice of Autonomy.* Cambridge University Press.

Easterlin, R. (2001). Income and Happiness: Towards a Unified Theory. *The Economic Journal 111*(473), 465–484.

Easterlin, R. (2003). Building a Better Theory of Well-Being. *IZA Discussion Paper No. 742.*

Ekstrom, L. (2005). Autonomy and Personal Integration. In J. Taylor (Ed.), *Personal autonomy: New essays on personal autonomy and its role in contemporary moral philosophy*, pp. 143–161. Cambridge University Press.

Ellerman, D. (2004). Autonomy-Respecting Assistance: Toward An Alternative Theory of Development Assistance. *Review of Social Economy 62*(2), 149–168.

Elster, J. (1983). *Sour Grapes: Studies in the Subversion of Rationality.* Cambridge University Press.

Fagerlin, A., B. Zikmund-Fisher, and P. Ubel (2005). Cure me Even if it Kills me: Preferences for Invasive Cancer Treatment. *Medical Decision Making 25*(6), 614.

Fehr, E. and K. Schmidt (1999). A Theory of Fairness, Competition and Cooperation. *The Quarterly Journal of Economics 114*, 817–868.

Forgas, J., K. Williams, and L. Wheeler (2003). *The Social Mind: Cognitive and Motivational Aspects of Interpersonal Behavior.* Cambridge University Press.

Fouarge, D. and R. Layte (2005). The Duration of Spells of Poverty in Europe. *Journal of Social Policy 34*(3), 1–20.

Friedman, M. (2003). *Autonomy, Gender, Politics.* Oxford University Press.

Galor, O. and J. Zeira (1993). Income Distribution and Macroeconomics. *Review of Economic Studies 60*(1), 35–52.

Geanakoplos, J., D. Pearce, and E. Stacchetti (1989). Psychological Games and Sequential Rationality. *Games and Economic Behavior 1*(1), 60–79.

Ghosal, S. (2006). Equilibrium with Incomplete Preferences: Non-convexity and Strategic Complementarity. *mimeo, University of Warwick*.

Gilboa, I. and D. Schmeidler (2001). *A Theory of Case-based Decisions*. Cambridge University Press.

Green, J. and D. Hojman (2007). Choice, Rationality and Welfare Measurement. *Harvard Institute of Economic Research Discussion Paper No. 2144*.

Hammond, P. (1976). Endogenous tastes and stable long-run choice. *Journal of Economic Theory 13*(2), 329–340.

Hardle, W. (1990). *Applied Nonparametric Regression*. Cambridge University Press.

Harsanyi, J. (1953). Welfare Economics of Variable Tastes. *Review of Economic Studies 21*, 204–213.

Hart, R. and D. Clark (1999). Does Size Matter? Exploring the Small Sample Properties of Maximum Likelihood Estimation. *Annual Meeting of the Midwest Political Science Association*, 15–17.

Hawton, K., P. Salkovskis, J. Kirk, and D. Clark (1989). *Cognitive Behaviour Therapy for Psychiatric Problems: A Practical Guide*. Oxford University Press, Oxford.

Heifetz, A. and E. Minelli (2006). Aspiration Traps. *mimeo*.

Hoff, K. and P. Pandey (2004). Belief Systems and Durable Inequalities: an Experimental Investigation of Indian Caste. *Policy Research Working Paper No 3351*.

Hoff, K. and A. Sen (2006). The Kin System as a Poverty Trap? In *Poverty Traps*, pp. 95–115. Princeton University Press.

Jalan, J. and M. Ravallion (1998). Determinants of transient and chronic poverty: evidence from rural China. *Policy Research Working Paper No 1936*.

Kahneman, D. (2000). Experienced Utility and Objective Happiness: A Moment-Based Approach. In D. Kahneman and A. Tversky (Eds.), *Choices, Values and Frames*, pp. 673–692. New York: Cambridge University Press.

Kahneman, D. and A. Tversky (1979). Prospect Theory: an Analysis of Decision under Risk. *Econometrica 47*(2), 263–291.

Kahneman, D., P. Wakker, and R. Sarin (1997). Back to Bentham? Explorations of Experienced Utility. *The Quarterly Journal of Economics 112*(2), 375–405.

Kannan, K. (2000). Poverty Alleviation as Advancing Basic Human Capabilities: Kerala's Achievements Compared. In G. Parayil (Ed.), *Kerala: The Development Experience*, pp. 40–65. New York. NY: Zed Books.

Karandikar, R., D. Mookherjee, D. Ray, and F. Vega-Redondo (1998). Evolving Aspirations and Cooperation. *Journal of Economic Theory 80*(2), 292–331.

Kasser, T. and R. Ryan (1993). A Dark Side of the American Dream: Correlates of Financial Success as a Central Life Aspiration. *J Pers Soc Psychol 65*(2), 410–22.

Koszegi, B. (2005). Utility from Anticipation and Personal Equilibrium. *University of California Berkeley, mimeo.*

Koszegi, B. and M. Rabin (2006). A Model of Reference-Dependent Preferences. *The Quarterly Journal of Economics 121*(4), 1133–1165.

Koszegi, B. and M. Rabin (2007). Reference-dependent Risk Attitudes. *American Economic Review 97*(4), 1047–1073.

Koszegi, B. and M. Rabin (2008). Choices, Situations, and Happiness. *Journal of Public Economics 92*(8-9), 1821–1832.

Kremer, M. (1993). The O-Ring Theory of Economic Development. *Quarterly Journal of Economics 108*(3), 551–575.

Lamm, H. and T. Schwinger (1980, dec). Norms Concerning Distributive Justice: Are Needs Taken Into Consideration in Allocation Decisions? *Social Psychology Quarterly 43*(4), 425–429.

Lazarus, R. and S. Folkman (1984). *Stress, Appraisal, and Coping.* Springer Pub Co.

Loewenstein, G. and P. Ubel (2008). Hedonic adaptation and the role of decision and experience utility in public policy. *Journal of Public Economics 92*(8-9), 1795–1810.

Loury, G. (1981). Intergenerational Transfers and the Distribution of Earnings. *Econometrica 49*(4), 843–867.

Mandler, M. (2005). Incomplete Preferences and Rational Intransitivity of Choice. *Games and Economic Behavior 50*(2), 255–277.

Manski, C. (1999). *Identification Problems in the Social Sciences*. Harvard University Press.

Manzini, P. and M. Mariotti (2009). Categorize Then Choose: Boundedly Rational Choice and Welfare. *mimeo, Queen Mary, University of London*.

Marmot, M. and R. Wilkinson (2006). *Social Determinants of Health*. Oxford University Press, USA.

Mas-Colell, A., M. Whinston, and J. Green (1995). *Microeconomic Theory*. Oxford University Press.

Maslow, A. (1943). A Theory of Human Motivation. *Psychological Review 50*, 370–396.

McGregor, D. (1960). *The Human Side of Enterprise*. New York: McGraw-Hill.

Mckinney, L. and L. Newman (2002). Anticipating Responses to One's Own Misdeeds: Repressive Coping and the Prediction of Others' Reactions to Inconsiderate Behavior. *Journal of Social and Clinical Psychology 21*(4), 427–440.

Milgrom, P. and C. Shannon (1994). Monotone Comparative Statics. *Econometrica 62*(1), 157–80.

Mookherjee, D. and D. Ray (2003). Persistent Inequality. *Review of Economic Studies 70*(2), 369–393.

Moreira, V. (2006). Poverty and Psychopathology. In *Understanding Poverty*. New York, US: Kluwer Academic/Plenun Publishers.

Mullainathan, S. (2005). Development Economics Through The Lens Of Psychology. In *Annual World Bank Conference on Development Economics: Lessons of Experience*. The World Bank and Oxford University Press.

Mullainathan, S. (2006). Better Choices to Reduce Poverty. In *Understanding Poverty*. Oxford, UK: Oxford University Press.

Murnighan, J. and A. Roth (2006). Some of the Ancient History of Experimental Economics and Social Psychology: Reminiscences and Analysis of a Fruitful Collaboration. In *Social Psychology and Economics*. New Jersey: Lawrence Erlbaum Assoc Inc.

Narayan, D. (2002). *Empowerment and Poverty Reduction: A Source-book*. Washington DC: World Bank Publications.

Nelson, R. (1956). A Theory of the Low-Level Equilibrium Trap in Underdeveloped Economies. *The American Economic Review 46*(5), 894–908.

Pollak, R. (1978). Endogenous Tastes in Demand and Welfare Analysis. *The American Economic Review 68*(2), 374–379.

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

Quah, D. (1993). Empirical Cross-section Dynamics in Economic Growth. *European Economic Review 37*(2-3), 426–434.

Rabin, M. (1993). Incorporating Fairness into Game Theory and Economics. *The American Economic Review 83*(5), 1281–1302.

Rao, V. and M. Walton (2004). Culture and Public Action: Relationality, Equality of Agency, and Development. In *Culture and Public Action*, pp. 3–36. The World Bank, Washington DC.

Ray, D. (2006). Aspirations, Poverty, and Economic Change. In A. Banerjee, R. Benabou, and D. Mookherjee (Eds.), *Understanding Poverty*. Oxford, UK: Oxford University Press.

Roberts, P. (2000). *Education, Literacy, and Humanization: Exploring the Work of Paulo Freire*. Greenwood Press.

Rubinstein, A. and Y. Salant (2008). (A,f) Choice with Frames. *The Review of Economic Studies 75*, 1287.

Ryan, R., V. Chirkov, T. Little, K. Sheldon, E. Timoshina, and E. Deci (1999). The American Dream in Russia: Extrinsic Aspirations and Well-Being in Two Cultures. *Personality and Social Psychology Bulletin 25*(12), 1509.

Ryan, R. and E. Deci (2006). Self-regulation and the Problem of Human Autonomy: Does Psychology need Choice, Self-determination, and Will? *Journal of Personality 74*(6), 1557–1586.

Schopler, J. and C. Insko (1992). The Discontinuity Effect in Interpersonal and Intergroup Relations: Generality and Mediation. *European Review of Social Psychology 3*(1), 121–151.

Schor, J. (1998). *The Overspent American: Upscaling, Downshifting, and the New Consumer*. Basic Books.

Selten, R. (1998). Aspiration Adaptation Theory. *Journal of Mathematical Psychology 42*(2-3), 191–214.

Sen, A. (1971). Choice Functions and Revealed Preference. *Review of Economic Studies 38*(3), 307–317.

Sen, A. (1977). Rational Fools: A Critique of the Behavioral Foundations of Economic Theory. *Philosophy & Public Affairs 6*(4), 317–344.

Sen, A. (1985). *Commodities and Capabilities*. North-Holland Amsterdam.

Sevdalis, N. and N. Harvey (2007). Biased forecasting of postdecisional affect. *Psychological Science 18*(8), 678–681.

Sewell, W. H. (1964). Community of residence and college plans. *American Sociological Review 29*(1), 24–38.

Sewell, W. H., A. O. Haller, and M. A. Straus (1957). Social status and educational and occupational aspiration. *American Sociological Review 22*(1), 67–73.

Sewell, W. H. and V. P. Shah (1968). Social class, parental encouragement, and educational aspirations. *The American Journal of Sociology 73*(5), 559–572.

Shafer, W. and H. Sonnenschein (1975). Equilibrium in abstract economies without ordered preferences. *Journal of Mathematical Economics 2*, 345–348.

Shalev, J. (2000). Loss aversion equilibrium. *International Journal of Game Theory 29*(2), 269–287.

Simon, H. (1955). A Behavioral Model of Rational Choice. *The Quarterly Journal of Economics 69*(1), 99–118.

Simon, H. (1959). Theories of Decision Making in Economics and Behavioural Science. *The American Economic Review 49*(1), 253–83.

Stark, O. (2006). Status Aspirations, Wealth Inequality, and Economic Growth. *Review of Development Economics 10*(1), 171–176.

Stark, O. and J. Taylor (1991). Migration Incentives, Migration Types: The Role of Relative Deprivation. *The Economic Journal 101*(408), 1163–1178.

Stern, N., J. Dethier, and F. Rogers (2005). *Growth and Empowerment: Making Development Happen*. Cambridge, Massachusetts and London, England: Munich Lectures in Economics, MIT Press.

Sugden, R. (2004). The Opportunity Criterion: Consumer Sovereignty Without the Assumption of Coherent Preferences. *American Economic Review 94*(4), 1014–1033.

Sunstein, C. and R. Thaler (2003). Libertarian paternalism. *American Economic Review 93*(2), 175–179.

Sunstein, C. and R. Thaler (2008). *Nudge: Improving Decisions About Health, Wealth, and Happiness*. New Haven, CT: Yale University Press.

Tarski, A. (1955). A Lattice Theoretical Fixed-point Theorem and its Applications. *Pacific J. of Math 5*, 285–309.

Traub, S. (1999). *Framing effects in taxation: an empirical study using the German income tax schedule*. Physica-Verlag, Heidelberg.

Tversky, A. and D. Kahneman (1991). Loss Aversion In Riskless Choice: A Reference-Dependent Model. *The Quarterly Journal of Economics 106*(4), 1039–1061.

Weizsäcker von, C. (1971). Notes on Endogenous Change of Tastes. *Journal of Economic Theory 3*, 345–372.

Wertenbroch, K. (1998). Consumption Self-control by Rationing Purchase Quantities of Virtue and Vice. *Marketing Science 4*(17), 317–337.

Williams, G., M. Gagne, A. Mushlin, and E. Deci (2005). Motivation for Behavior Change in Patients with Chest Pain. *Health Education 105*(4), 304–321.

Wilson, W. (1987). *The Truly Disadvantaged*. University of Chicago Press.

Wit, A. and N. Kerr (2002). Me versus just us versus us all" categorization and cooperation in nested social dilemmas. *Journal of Personality and Social Psychology 83*(3), 616–637.

Wolfe, R. (1997, September). Omodel: Stata modules to perform tests on ordered probit and ordered logit models. *Statistical Software Components, Boston College Department of Economics*.

Woo, H. (1984). *The Unseen Dimensions of Wealth*. Victoria Press.

Yamagishi, T. and T. Kiyonari (2000). The group as the container of generalized reciprocity. *Social Psychology Quarterly 63*(2), 116–132.

Yuki, M., W. Maddux, M. Brewer, and K. Takemura (2005). Cross-cultural Differences in Relationship and Group-based Trust. *Personality and Social Psychology Bulletin 31*(1), 48.

Zizzo, D. and A. Oswald (2001). Are People Willing to Pay to Reduce Others' Incomes? *Annales d'économie et de Statistique 63–64*, 39–65.