Incentives, social preferences and pro-social behavior: an experimental approach

A Dissertation Presented
by

Sandra Polania-Reyes

Supervisor**: Prof. Luigi Luini**

Thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Economics
University of Siena

June 2015

Contents

Acknowledgements

This dissertation would have been impossible to complete without the help, patience and guidance of my chair and advisor, Professor Luigi Luini.

I want to thank my mentor and friend Samuel Bowles. There are no words to express how grateful I am for his encouragement, the many hours of talks and discussions and the valuable emails. I am grateful for being a direct witness of his example and grateful for his life. Sam, thank you for your patience and your confidence in my capabilities especially at moments when I was not able to see them.

I am also in debt to my co-author and coach, Professor Orazio Attanasio, for his guidance and confidence in my work, his useful comments and ideas. I thank especially Professor Syngjoo Choi for his support and challenging questions.

I thank Juan Camilo Cárdenas for his guidance, great feedback and special care. I am in debt to Ana Maria Roldán, Adriana Molina and the team of research assistants that made the data collection possible. The fieldwork for this project was possible in part due to the generous support from the Development Bank of Latin America (CAF), the University of Los Andes, the Javeriana University and the McCarthur Foundation. I want to thank to my colleague and friend Patricia Padilla for her dedication to our cause and her invaluable company during the fieldwork. Thanks also to Miguel Espinosa, Vivian Rodriguez and the team of surveyors in Cartagena, Colombia.

I want to express all my gratitude to the academic community in London and especially to all the faculty members, the staff and students at the Department of Economics at the University of Siena from whom I learned so much through my time there. I have fond memories of the long days and nights, the discussions, the encouragement and the coffee breaks. I also want to thank my Colombian friends and colleagues; their support and joy have inspired my work.

My time as student of University of Siena has been the most challenging and enriching in all spheres of my life. I discovered human imperfection and divine grace. It was an amazing process of learning and growing, but most of all one of support and encouragement. I chose U Siena with clear expectations in the academic, professional and spiritual arena. I was not disappointed but overwhelmed and joyful.

This research would have been impossible without the participation of the beneficiaries of Familias in Acción in Cartagena, the students from Javeriana University and the Common Pool Resource users in rural areas in Colombia. My inspiration comes from interacting with others, my motivation to contribute will last thanks to them.

I thank to my friends back home for all their support. To my family and my family-in-law, for their patience and prayers. I would not be here now if it is not for my husband, lover, sponsor, coach and co-author, David Echeverry. Your support, devotion, encouragement and endless and enriching discussions made this work invaluable to me.

My daughter Cristina brought me the gift of motherhood and the delicious complexities that make life harder but meaningful and joyful. I dedicate this achievement to you both.

Introduction

This dissertation is a study of how prosocial behavior is enhanced by economic incentives once the heterogeneity of social preferences among the community is taken into account. The objectives are to develop an empirical characterization of how individual behavior changes under different incentives, and to further understand the effects of different institutions to promote prosocial behavior. In order to design policy interventions that promote the reach of more efficient economic outcomes, we need to understand how individual and group behavior is rooted in heterogenous motivations. This has always been of interest to policy makers, lawyers and economists and it has recently become more prominent with the growing interest in social capital, trust, altruism and the like, as essential to the working of modern economies. The results are organized into three chapters.

The first chapter examines the sometimes surprising, sometimes perverse effects of economic incentives when they interact with social preferences or intrinsic motives. This study will be what we would call "second generation" behavioral economics. Having focused on the psychological assumptions of economics and having found them to be not entirely adequate, it's time to figure out what difference this makes for standard questions in economics -like mechanism design, public economics, etc[1].

We address the fact that interventions by policy makers and mechanism designers may go seriously wrong when the effect of incentives on preferences is ignored. The problem is based on the fact that policy outcomes may deviate from those predicted if preferences, which are assumed to be exogenous when designing the policy, depend on the incentives or other policy instruments used. At least informally, this possibility has been recognized since Titmuss' book on blood donations a generation ago. Empirical evidence has been lacking; partly for that reason, theoretical public economics has not addressed the problem of motivational crowding out.

The main recommendation of the chapter is to identify the conditions under which crowding out (or the opposite, i.e. crowding in) occurs, and to use this information to suggest new approaches to incentives and public policy. Some of these implications are not what one would expect. For example, in the presence of crowding in the sophisticated mechanism designer who is aware of the problem will sometimes make greater use of explicit incentives rather than less, as the critics of incentives generally assume.

The second chapter accounts for heterogeneous social preferences and explore possible determinants of these preferences. This is important to address effectiveness and efficacy of incentives to promote pro-social behavior. Our approach contributes to recent literature that we would call "third generation" behavioral economics. On one hand, we propose a structural estimation of theoretical parameters of social preferences in order to identify social preferences distributions among the population. On the other hand, we aim to use our estimation in order to evaluate efficiency among a class of incentives.

We present the results from a Common Pool Resource (CPR) game with subjects from different communities in Colombia. First, we drop the assumption that all individuals are all self-regarding and develop several models of pure Nash strategies for our CPR game when

[1] See Datta & Mullainathan (2014).

individuals are motivated by a combination of self-interest and preferences for altruism, reciprocity or inequity aversion. Second, we estimate individual heterogeneity by using a random coefficients model and classify individual social preferences according to their behavior in the baseline phase by assigning a type to every participant. Third, we compare the role of heterogeneity of preferences in social efficiency across incentives and confirm the existence of different effects of incentives on each type and find the subsidy as the most socially efficient incentive. Finally, we obtain *exogenous* determinants of individual type such as socio-economic characteristics, perceptions on the CPR, perceived interest in cooperation among the community, whether the participant does volunteer work and whether the CPR is the household main economic activity of the household; we also obtain *endogenous* determinants such as the composition of types in the group and their demographic characteristics.

The third chapter examines how experimental data can shed light on a topic in development economics such as an intervention design to reduce poverty. So the main message of the paper will be that policy interventions may have collateral positive effects different from the target areas of the intervention. This possibility has been mentioned in the literature, but empirical evidence on it has been lacking. We present empirical evidence suggesting that interventions such a Conditional Cash Transfer (CCT) Program may affect the communities' cooperative behavior.

We use a field experiment in order to measure the effect of a CCT on social capital. This collateral effect of interventions has recently become more prominent to policy makers. In particular, this suggests new considerations to incentives and public policy as incentives may change not only targeted desired behavior but alter the social relationships among the actors, and the preexisting normative frameworks of the actors.

We implement a public goods game; the experimental data is complemented by a collection of information on networks among players as well as the traditional survey measures to measure social capital. We present a difference-in-difference estimation allowing us to identify the effect of the program on cooperation in the CCT Program. Our sample is the largest used in the experimental literature on cooperation: 1295 poor households in Cartagena, Colombia. This econometric model and the applications to public policy interventions and to the study of social preferences in the field are entirely novel.

I. Economic incentives and social preferences: substitutes or complements?

This chapter is joint work with Sam Bowles.[2] Another version of this chapter has been published as Working Paper N. 617 (http://www.econ-pol.unisi.it/dipartimento/it/node/1385) and is forthcoming at Journal of Economic Literature.

Abstract

Explicit economic incentives designed to increase contributions to public goods and to promote other pro-social behavior sometimes are counterproductive or less effective than would be predicted among entirely self-interested individuals. This may occur when incentives adversely affect individuals' altruism, ethical norms, intrinsic motives to serve the public, and other social preferences. In the 50 experimental studies that we survey these effects are common, so that incentives and social preferences may be either substitutes (crowding out) or complements. We provide evidence for four mechanisms that may account for these incentive effects on preferences, based on the fact that incentives may (i) provide information about the person who implemented the incentive, (ii) frame the decision situation so as to suggest appropriate behavior, (iii) compromise a control averse individual's sense of autonomy and (iv) affect the process by which people learn new preferences. An implication of the fact that incentives affect preferences is that the evaluation of public policy must be restricted to allocations that are supportable as Nash equilibria when account is taken of these crowding effects. We show that well designed fines, subsidies and the like minimize crowding out and may even do the opposite, making incentives and social preferences complements rather than substitutes.

JEL codes: D02 (Institutions); D03 (Behavioral Economics; Underlying Principles); D04 (Microeconomic Policy: Formulation; Implementation; Evaluation); D83 (Search; Learning; Information and Knowledge; Communication; Belief); E61 (Policy Objectives; Policy Designs and Consistency); H41 (Public goods); Z13 (Social norms and social capital)

Keywords: Behavioral experiments, social preferences, endogenous preferences, motivational crowding, explicit incentives.

[2] For their contributions to this paper we thank Mahzarin Banaji, Abigail Barr, Iris Bohnet, Stephen Burks, Juan Camilo Cárdenas, Jeff Carpenter, David Echeverry, Ernst Fehr, Urs Fischbacher, Christina Fong, Simon Gächter, Roberto Galbiati, Joshua Greene, Sung-Ha Hwang, Jonathan Haidt, Daniel Houser, Steffen Huck, Bernd Irlenbusch, Magnus Johannesson, Esther Kessler, Nicola Lacetera, Maria Claudia López, Thomas Schelling, Carlos Sickert-Rodriguez, David Skuse, Joel Sobel, John Stranlund, Erte Xiao, Elisabeth Wood as well as three anonymous referees and the editor. Financial support by the Behavioral Sciences Program of the Santa Fe Institute, the U.S. National Science Foundation, the University of Siena and the European Science Foundation is gratefully acknowledged.

1. Introduction

Thomas Schelling recalls his "exciting and stimulating times" in the early 1950s as a young staffer in the Executive Office of the President. "People worked long hours," he remembered in a recent communication to one of us, "and felt compensated by the sense of accomplishment and ... personal importance. Regularly a Friday afternoon meeting at the White House would go on until 8 or 9, when the chairman would suggest resuming Saturday morning. Nobody demurred. We all knew it was important, and we were important. ... What happened when the President issued an order that anyone who worked on Saturday was to receive overtime pay…? Saturday meetings virtually disappeared."

Was Schelling's experience atypical? Incentives work, often affecting the targeted behavior almost exactly as conventional economic theory predicts: textbook examples include the work response of Tunisian sharecroppers and American windshield installers, and the reduced criminal activities of former Italian convicts who could expect more severe sentences if convicted (Laffont and Matoussi (1995), Lazear (2000), Drago, Galbiati and Vertova (2009)).

But explicit economic incentives sometimes have surprisingly limited effects and may even be counterproductive. Substantial rewards for high school matriculation in a randomized experiment in Israel had no impact on boys and little effect on girls except among those already quite likely to matriculate (Angrist and Lavy (2009)). Large and in most cases immediate cash payment in return for tested scholastic achievement in 250 urban schools in the U.S. were almost entirely ineffective, while incentives for student inputs (reading a book, for example) had the intended, if modest effects (Fryer (2011)). In an unusual natural experiment, the imposition of fines designed to reduce hospital stays in Norway had the opposite effect (Holmås, Kjerstad, Lurås, et al. (2010)) while in England hospital stays were greatly reduced by a policy designed to evoke shame and pride in hospital managers rather than the calculus of profit and loss (Besley, Bevan and Burchardi (2009)).

Anecdotal accounts of what appear to be even more dramatic cases of counterproductive incentives are common. On December 1, 2001 the Boston Fire Department terminated its policy of unlimited paid sick days, replacing it with a 15-day sick day limit; pay would be docked for firemen exceeding the limit. The firemen responded to the new incentives: those calling in sick on Christmas and New Year's Day increased tenfold over the previous year. The Fire Commissioner retaliated by cancelling their holiday bonus checks (Belkin (2002)). The firemen were unimpressed: the year following they claimed 13,431 sick days; up from 6,432 the previous year (Greenberger (2003)). Many of the firemen, apparently angered by the new system, abused it or abandoned their previous ethic of serving the public even when injured or not feeling well.

Not surprisingly, then, since Richard Titmuss' *The Gift Relationship: From Blood Donations to Social Policy*, economists have been intrigued by the claim that policies based on explicit economic incentives may be counterproductive when they induce people to adopt what Titmuss called a 'market mentality' or in some other way compromise pre-existing values to act in socially beneficial ways. But few were persuaded (Solow (1971), Arrow (1972), Bliss (1972)).

At the time of its publication there were two strong reasons to doubt Titmuss' claim: there was little hard evidence that social preferences are important influences on individual

behavior; and there was even less evidence (in the Titmuss (1971) book or elsewhere) that social preferences would be undermined by explicit economic incentives (which we will call simply "incentives" without the adjectives, meaning interventions to influence behavior by altering the economic costs or benefits of some targeted activity.)

Theoretical and empirical advances over the intervening years provide the basis for a reconsideration of these issues (Kreps (1997), Elster (1998), Rabin (1998), Loewenstein (2000), Sobel (2002); surveyed in Meier (2007a)). First, evidence from both the behavioral experimental laboratory and the field is consistent with the view that social preferences are important influences on economic behavior (Fehr, Gächter and Kirchsteiger (1997), Bewley (1999), Fehr and Schmidt (1999), Fehr and Gächter (2000), Young and Burke (2001), Bandiera, Barankay and Rasul (2005), Falk, Fehr and Fischbacher (2005), Fehr, Klein and Schmidt (2007), DellaVigna (2009), Leider, Möbius, Rosenblat, et al. (2009), Sloof and Sonnemans (2011)).

Second, the importance of incomplete contracts has been widely recognized in theoretical works and studied empirically (Stiglitz (1987), Laffont and Matoussi (1995), Tirole (1999)). Partly as a result, the terms trust, reciprocity, fairness, gift exchange and social capital now appear in the modeling and empirical study of principal-agent relationships, the provision of public goods, and other standard economic applications, often referring to the social norms and identities that underwrite mutually beneficial exchange in the absence of complete contracts (Arrow (1971), Becker (1976), Akerlof (1984), Helsley and Strange (2000), Benabou and Tirole (2006), MacLeod (2007), Sliwka (2007), Ellingsen and Johannesson (2008), Akerlof and Kranton (2010)).

Third , advances in the theory of public policy have addressed cases in which incentives affect both beliefs and preferences and may thus have unintended effects (Lucas (1976), Taylor (1987), Bowles (1989), Aaron (1994), Sunstein (1996), Frey (1997), Cooter (2000), Bowles (2004), Bar-Gill and Fershtman (2005), Sobel (2005), Cervellati, Esteban and Kranich (2010)).

Finally, a substantial literature (surveyed in Frey and Jegen (2001), Bowles (2008), Gneezy, Meier and Rey-Biel (2011)) emerged showing that explicit incentives are sometimes counterproductive.

2. Overview: Incentives and social preferences as substitutes or complements

We use the term "social preferences" to refer to motives such as altruism, reciprocity, intrinsic pleasure in helping others, inequity aversion, ethical commitments and other motives that induce people to help others more than would an own-material-payoff maximizing individual. Our use of the term is thus not restricted to cases in which the actor assigns some value to the payoffs received by another person, as in the utility functions of Fehr and Schmidt (1999), Rabin (1993) and Levine (1998). While these functions provide a convenient way to model some of the motivations for pro-social behavior, we use the broader definition because moral, intrinsic, or other reasons unrelated to a concern for another's payoffs often motivate people to help others, adhere to social norms, and act in other pro-social ways even when it is personally costly to do so. A person, for example, may adhere to a social norm not because of the harm that a transgression would do to others, but because of the kind of person

she would like to be; helping the homeless may be motivated by Andreoni's "warm glow" of giving rather than a concern with the wellbeing of the poor (Andreoni (1990)).

The standard (if generally implicit) assumption in economics is that the behavioral functions relevant for mechanism design, public economics and related fields are separable in social preferences (should they exist) and incentives. This means, for example, that the citizen's response to variations in a subsidy for contributions to a public good is independent of her pre-existing level of social preferences. It also means that the effect of variations in her pre-existing non-economic motivations on the citizen's level of contributions does not depend on the presence or magnitude of incentives.

We call this the separability assumption. (A recent example Levitt and List (2007).) It is a useful simplification for problems in which social preferences can safely be ignored, as in the choice of apples versus oranges and other standard textbook examples. But it can also be misleading, as it implies that taxes, subsidies, and other incentives affect behavior only by altering the economic costs and benefits of the targeted activities. We will present evidence that the separability assumption often does not hold, and as a result that social preferences may be either heightened by incentives appealing to self-interest or, the more commonly observed case, affected adversely by incentives.

This is illustrated in Figure 1 where, due to the effect of incentives on preferences, the total – direct and indirect -- effect of the incentive may fall short of that which works directly on the costs and benefits of the targeted activity. In this case we say that incentives crowd out social preferences and that incentives and social preferences are substitutes: the effect of each on the targeted activity declines, the greater is the level of the other. Where the effect on social preferences is positive, crowding in occurs and social preferences and incentives are complements, the level of each enhancing the effect of the other.

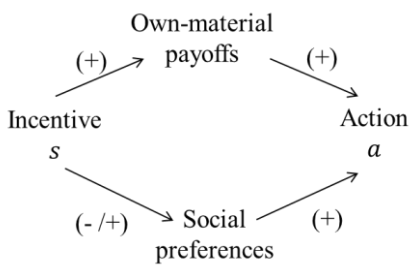


Figure 1. Crowding effects of incentives: The direct and indirect effects of incentives on contribution to a public good (a) The effect of an incentive (s) on social preferences may be either to reduce their behavioral salience for the action (social preferences are state-dependent) or to affect the manner in which preferences are updated, thereby altering the individual's social preferences (endogenous social preferences). Crowding out occurs when the effect of an incentive on social preferences is negative (assuming that the effect of social preferences on the action is positive, as shown). Crowding in (the opposite) also occurs.

The possibility that incentives designed for material payoff-maximizers might have adverse effects is a familiar theme in political science (Taylor (1987), Grant (2011)), psychology (Deci (1975)), sociology (Healy (2006)), and the other social sciences; but it has found few adherents in economics. The reason is that we have adopted a simplifying strategy that goes back at least to John Stuart Mill (1867[1848]): 97)

> [Political economy] does not treat of the whole of man's nature... it is concerned with him solely as a being who desires to possess wealth,… it predicts only such ...phenomena ...as take place in consequence of the pursuit of wealth. It makes entire abstraction of every other human passion or motive.

In other words, we ignore the two lower arrows in Figure 1. But recent experimental and other evidence has prompted many economists to reconsider Mill's simplification. To further

this reconsideration we here provide a taxonomy of incentive effects on preferences based on two distinctions: their nature and their causes.

Concerning the first, people often react to the mere presence of incentives rather than their extent (Gneezy (2003)): giving to charity when tax breaks are involved (whatever their magnitude) may feel different or send a different signal than would be the case in the absence of these incentives. But the extent of the incentive may also matter. Thus the effects of incentives on social preferences may be either marginal (depending continuously on the level of the incentive) or categorical (the presence of incentives affecting social preferences independently of their level) or a combination of the two.

We also distinguish between two causes of incentive effects on preferences. First, behavior is acutely sensitive to the nature of the decision situation (Ross and Nisbett (1991), Tversky and Kahneman (1981)); and, as we will see, the presence or extent of incentives provides information about the situation. A psychologist might say that preferences are "situation-dependent" and that incentives provide situational clues. We say that the preferences are state-dependent, with differing incentives constituting different states. In the next section we offer a model of incentive-state-dependent preferences and provide data indicating that both categorical and marginal crowding out occurs.

State-dependence arises because actions are motivated by a heterogeneous repertoire of preferences –from spiteful to payoff-maximizing to generous, for example -- the salience of which depends on the nature of the decision situation – interacting with a domineering supervisor, shopping, or relating to one's neighbors, for example. To see how this works, think about gifts. Economists know that money is the perfect gift – it replaces the giver's less well-informed choice of a present by the recipient's own choice. But when the holidays come around few economists give money to their friends, family and colleagues. This is because we also know that money cannot convey thoughtfulness, concern, whimsy, or any of the other messages that non-monetary gifts sometimes express. A gift, we know, is more than a transfer of resources; it is a signal about the giver and her relationship to the recipient, and money changes the signal.

Can the same be said of incentives? A long tradition in psychology has concluded that it can:

> The multiple meanings of ... tangible rewards are reflected in our everyday distinction among bribes and bonuses, incentives and salaries. … they carry different connotations concerning, for example, (i) the likely conditions under which the reward was offered, (ii) the presumed motives of the person administering the reward, and (iii) the relationship between the agent and the recipient of the reward (Lepper, Sagotsky, Dafoe, et al. (1982) numbers added).

As Mark Lepper and his coauthors say, incentives may affect preferences for a reason familiar to economists, that is because they indicate "the presumed motives of the person administering the reward." By implementing an incentive, a principal reveals information about his or her intentions (own payoff maximizing vs. fair-minded, for example) as well as beliefs about the target of the incentives (hardworking or not) and the targeted behavior (how onerous it is, for example.) This information, in turn, may then affect the target's non-economic motivation to undertake the task at hand. In section 5 we present experimental evidence that the information provided about the principal can sometimes attenuate or even reverse the intended effect of the incentive. Of course when an incentive provides good news about the principal's intentions or type—when rewards are offered, for example, rather than

fines– it may recruit the target's social preferences to work synergistically with the direct effect of the incentive In this case incentives and social preferences become complements rather than substitutes. We will see (in section 9 and 11) that this crowding in phenomenon is sometimes observed in experiments, for example, when the principals implementing incentives are peers in a public goods game who pay to fine free riders in order to support cooperative norms.

But there are other reasons, less familiar to economists, for state-dependence: reasons that do not concern information about the principal, and that may be at work even in non-strategic settings. A second mechanism is that incentives provide cues about (as Lepper and his coauthors put it) "the likely conditions under which the reward was offered:" by framing a decision situation, economic incentives may provide cues for appropriate behavior. This second mechanism is distinguished from the first in the experimental evidence by the fact that in the former the incentives are implemented by a principal who is a player in the game; while in the latter the targets of the incentive are not playing against the incentive designer; rather the incentives are introduced by the experimenter.

Situational cues may be very subtle, and our responses to them unwitting. When experimental subjects had the opportunity to cheat on a test and as a result to gain higher monetary rewards, less than quarter did so when the room was brightly lit, but more than half cheated when the room was slightly less well lit (the variations in lighting had no effect on the observability of cheating.) In another experiment subjects who wore (nonprescription) dark glasses were much less generous to their partner in a Dictator Game than were those outfitted with clear glasses (Zhong, Gino and Bohns (2010)). The dark glasses and darkened room gave the subjects a sense of anonymity, the researchers found. But it was entirely illusory: it is difficult to imagine that a subject could really think that his own wearing dark glasses would make him less observable, especially given that the experiment was conducted at computer terminals in closed cubicles.

The degree of anonymity differs dramatically as we move between family, workplace, marketplace and other domains of social interaction. Fiske (1992) provides a taxonomy of four psychological models corresponding to distinct kinds of social relationships: authoritarian, communal, egalitarian and market, each with culturally prescribed patterns of appropriate behavior. Depending on the information they convey, incentives may signal that the situation corresponds to any one of these four types, and therefore evoke distinctive responses.

We will see that a plausible explanation of some of the framing effects of incentives observed in experiments is that it occurs because market-like incentives trigger what psychologists term "moral disengagement" (Bandura (1991), a process that occurs because "people can switch their ethicality on and off" (Shu, Gino and Bazerman (2009):31). In section 6 we review experiments in which crowding out appears to have been the result of moral disengagement. Depending on the information they convey, incentives may also trigger the opposite – moral engagement – and, as we will see in section 9, experiments provide a few examples of this form of crowding in, illustrating the possible synergy or complementarity between social preferences and incentives.

The third mechanism that makes social preferences state-dependent is the crowding out of intrinsic motives by incentives (or constraints) that compromise a subject's sense of autonomy (Deci and Ryan (1985), Deci, Koestner and Ryan (1999)). These effects may occur

in strategic situations where the bad news that incentives convey concerns the desire of a principal to control the agent. But most of the experimental evidence for this third crowding out mechanism comes from non-strategic settings (the experimenters, not a principal implements the incentive.) The underlying psychological mechanism appears to be a fundamental desire for "feelings of competence and self-determination" that are associated with intrinsically motivated behavior (Deci (1975)).

According to this interpretation, where people derive pleasure from an action per se in the absence of other rewards, the introduction of incentives may 'over-justify' the activity and reduce the individual's sense of autonomy. This self-determination mechanism differs from the previous two mechanisms -- bad news about a principal and moral disengagement– because it arises from the target's desire for autonomy and does not depend on the target inferring negative information about a principal or clues about appropriate behavior. This is particularly evident in some early "over-justification" experiments in which when a financial reward was offered by the experimenter, children often forsook previously uncompensated activities in which they had enthusiastically engaged, like painting. More recent experiments show the same negative effects of incentives on altruistic behavior (Warneken and Tomasello (2008)). In the absence of rewards kids less than two years old avidly helped an adult retrieve an out of reach object; but after being rewarded with a toy for their helping behavior the helping rate fell by forty percent.

The fact that the incentive was a reward rather than a penalty suggests that it did not convey negative information about the incentive designer, but instead altered the meaning of the activity itself from one that expressed autonomy to one that expressed compliance. The interpretation that self-determination is involved in the negative response to incentives is consistent with the fact that close supervision or arbitrary temporal deadlines for completion of an otherwise enjoyable activity have effects very similar to financial rewards (Lepper, et al. (1982)). In section 7 we survey experimental evidence for this "control aversion" mechanism for state-dependent preferences.

We have just described three (partially overlapping) reasons why the state-dependent nature of preferences might lead to crowding out. For ease of reference we will call them "bad news," "moral disengagement," and "control aversion." But in addition to incentives altering the preferences that motivate an individual's action by altering the subject's sense of the situation, there is a second and quite different way that incentives may affect preferences. The type and extent of a society's use of economic incentives also may affect the process of preference-updating by which individuals acquire new tastes or social norms that will persist over long periods. Models from biology, anthropology, and economics allow us to formalize this learning process (Cavalli-Sforza and Feldman (1981), Boyd and Richerson (1985), Guth and Kliemt (1994), Bowles (1998), Bisin and Verdier (2011), Bowles and Gintis (2011)).

The key difference between endogenous and state-dependent preferences is that in the former case the effect of the incentive on preferences persists in the long run because the updating process on which cultural transmission is based typically occurs during youth and its effect endures over decades if not entire lifetimes. We say that incentives affect preferences in both the state-dependent and endogenous preference case, but the mechanism of the effect is different: in the former case the incentive is a reversible signal about the principal or the situation, in the latter the incentive alters the preference-updating process.

An example unrelated to incentives may clarify the difference between endogenous and state-dependent preferences. As Italian residents, your authors now eat a lot more pasta than we did in our countries of origin. Abstracting from possible international price differences, this could be another case of "when in Rome, do as the Romans." Or it might be that we have newly come to enjoy the taste of pasta, perhaps through extensive exposure to it while in Italy. Which case it is – state-dependent or endogenous preferences – would be revealed by what we will eat back in Bogotá or Santa Fe. If we go back to *arepas* or potatoes, then our taste for pasta was state-dependent. If we remain *pastaphiles*, then our preferences have endogenously changed.

Preferences may be endogenous in this sense because the extent to which a society relies on economic incentives – as opposed to other kinds of motivations and controls – may affect how people learn new preferences (evidence for the endogeneity of preferences is surveyed in Bowles (1998), (2004), Bowles and Gintis (2011).) The learning on which preference endogeneity is based is of course a long term process unlikely to be observed in a brief experiment. Nonetheless experiments may provide clues that learning is affected: we take as evidence consistent with preference endogeneity those cases in which crowding effects of incentives persist after the removal of the incentive. (Other explanations not involving endogenous preferences are generally also possible in these situations.)

In sections 3 and 4 we make explicit the underlying causal mechanisms through the use of models of state-dependent and endogenous preference formation, Table 1 summarizes the differences.

Source and Characterization (modeled in section §)	**Mechanisms**	**Description (§: section with empirical evidence relevant to this mechanism)**
State-dependent preferences Incentives affect the *behavioral salience of an individual's social preferences*, §3	**Information "bad news"**	Incentive signals the designer's type or beliefs about the target or the nature of the targeted task, and may convey illegitimate pursuit of self-interest by principal. §5
	Framing "moral disengagement"	Incentive signals the type of situation and hence appropriate behavior for the target, and may activate own payoff-maximizing modes of thought. §6
	Self-determination "control aversion"	Incentive affects target's sense of autonomy, and may signal unacceptable control and motivate resistance. §7
Endogenous preferences Incentives affect the *environment in which preferences are learned* and therefore the stationary *distribution of preference types in the population* (i.e. the fraction of population with social preferences), §4	**Conformist preference-updating**	Incentives reduce the perceived population fraction of social preference types. The extent to which a society relies on economic incentives – as opposed to other kinds of motivations and controls – will affect how people learn new preferences that may persist over long periods. §8

Table 1. Economic incentives and social preferences: Endogenous and state-dependent effects and mechanisms. As a result of the mechanisms listed incentives and social preferences may be either complements (crowding in) or substitutes (crowding out). In the conclusion we consider cases in which the degree of endogenous or state-dependent non-separability is subject to public policy because the crowding parameters $\lambda_m, \lambda_c, \Lambda_m$ and/or Λ_c may themselves be affected by incentives. Additional mechanisms for endogenous crowding out are provided in Bowles (2004).

Our empirical strategy (based on experimental results) is to observe the total effect of incentives on behavior and to note whether this differs from the predicted direct effect (the top arrows in Figure 1) in order to infer the effects of incentives on (unobserved) social

preferences and thereby on actions (the bottom two arrows). Our data set includes all the economic experiments we have been able to locate that allow a test of the separability assumption. Our tables include more than a hundred different subject pools, over twenty-six thousand subjects from 36 countries, playing Dictator, Trust, Ultimatum, Public Goods, Third Party Punishment, Common Pool Resource, Gift Exchange and other principal-agent games. These are all settings in which one's actions affect the payoffs of others so that social preferences may affect a subject's experimental behavior. We find evidence of nonseparability in all of these games. Because nonseparability, as we will see, arises from the social relationships among those imposing incentives and their targets and the nature of the incentive, and because game structures differ in this respect, it would be surprising if the nature and degree of non separability did not differ across these games. However, lacking a metric for non-separability that is comparable across games, we have not explored this possibility.

Few experiments have thus far been designed to address the causes of non-separability, so the inferences that we draw must be provisional. The experimental methods that have become standard in economics include playing for real stakes, excluding deception, and making explicit use of game theoretic concepts to clarify the role of incentives. As experimental methods differ considerably across disciplines, and for reasons of space, we limit the entries in the tables to experiments done by economists. We refer to a number of important experiments done using other methods in the text. All reported results are statistically significant at conventional levels unless noted.

Incentives may have counter-intuitive and counterproductive effects for reasons other than non-separability (Seabright (2009)). Strong monetary incentives, for example, may over-motivate an agent leading to greater than the optimal level of arousal. This appears to be the mechanism underlying the negative effects of high incentives found in three experiments by Ariely, Gneezy, Loewenstein, et al. (2009). In other cases incentives alter individuals' beliefs about the actions of others, with possibly counter-intended effects. This is often thought to be the case when tax authorities announce stiff penalties for underpayment, unwittingly letting the public know that that cheating is common and thereby promoting rather than deterring it. We do not consider these and other cases of counterproductive incentives where the mechanisms are unrelated to the non-separability of incentives and social preferences, which is the focus of this paper.

Some of the experimental results presented below may be explained by more than one of our four mechanisms that account for non-separability, either because the mechanisms are not mutually exclusive so that multiple mechanisms are at work, or because the experiment does not provide sufficient information to say which mechanism accounts for the evidence of non-separability. In these ambiguous cases we classified the experiment as an illustration of the mechanism which we thought best accounted for the crowding result that we report (We indicate in each table where such ambiguities occur and which other mechanisms may have been at work). As a preview, Figure 2 presents a summary of our findings, the size of the ellipses indicating the total number of studies that exhibit each of the four crowding out mechanisms in question, and the intersections giving the cases where multiple mechanisms may be involved.

There are two reasons why despite the considerable number of experiments in which preference effects of incentives appear to be at work it is difficult to estimate how prevalent these effects are in real economies. First, the experimental games involved are about social

dilemmas or sharing with others, that is, settings in which social preferences are likely to be important and therefore there is something to be crowded out or in. While the experimental evidence suggests that crowding out may affect blood donations or participation in community service projects, it does not have much to say about the effect of incentives on shopping behavior or cleaning hotel rooms. Second, while section 10 presents evidence that experimental play in these social dilemmas predicts behavior in some non-experimental situations, isolating social preferences from other influences on behavior in natural settings is difficult. We conclude in sections 11 and 12 with policy implications.

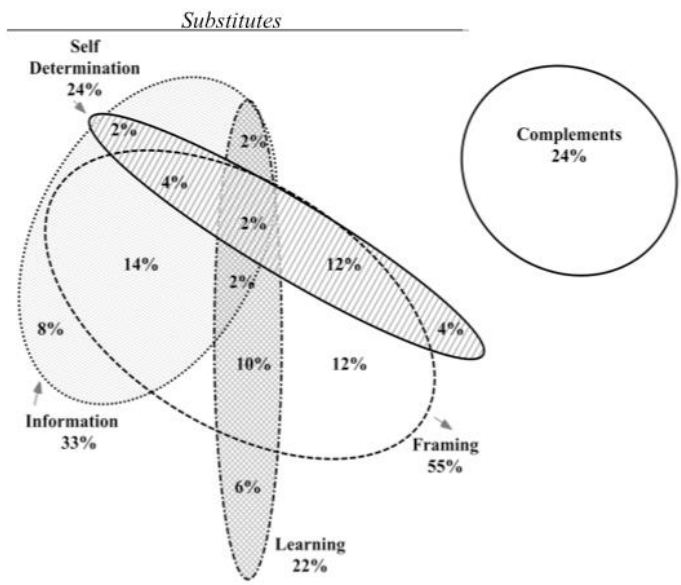


Figure 2. Summary of experimental evidence on the four crowding out mechanisms and crowding in. In the figure on the left the mechanisms accounting for crowding out are shown. The intersections show cases in which more than one mechanism may be involved. For example 14% of the experiments are consistent with both the framing and information about the incentive designer mechanisms. The circle in the upper right refers to crowding in (we have separated out the mechanisms in this case.) The numbers indicate the percentage of the total of 50 studies that exhibit the mechanisms indicated. There are no studies in the intersections that are blank.

3. Incentives as signals: a model of state-dependent preferences

In this section (following Hwang and Bowles (2011a) and Bowles and Hwang (2008)) we model incentive effects on state-dependent preferences and clarify the distinction between categorical and marginal incentive effects by means of an empirical illustration. We consider an individual who may bear a cost to take an action that confers benefits on others, which may be encouraged by a subsidy implemented by a social planner. Citizens also have values that may motivate such pro-social actions even in the absence of the subsidy. We study a single member of a community of identical citizens who may contribute to a public project by taking an action a at a cost $g(a)$ that is increasing and convex in its argument, and that may be offset partially by a subsidy s, that is proportional to the individual's level of contribution. The output of the project is available in equal measure to all, and it varies positively and linearly with A, the sum of the n members' contributions, according to φA where φ is a positive constant.

We express the individual's social preferences as v, the effect of an increase in the contribution level on the individual's utility that is unrelated to material payoffs. Thus we have the individual's utility

(1) $$u = j\ A - g(a) + as + av$$

we make explicit the sources of non-separability by the value function:

(2) $$v\left(s;\lambda_0,\lambda_c,\lambda_m\right) = \lambda_0\left(1 + 1\{s>0\}\lambda_c + s\lambda_m\right)$$

where the indicator $1\{s>0\}=1$ if $s>0$ and zero otherwise. In equation (2) $\lambda_0 \geq 0$ measures the citizen's baseline social preferences namely the citizens values in the absence of a subsidy, λ_c (which may be of either sign) measures the categorical effect of the presence of an incentive, and λ_m (which also may be of either sign) measures the marginal effect of variations in s on values for $s>0$. The crowding effects represented by λ_c and λ_m in (2) may arise because of any of the three mechanisms by which state-dependent preferences arise: bad news, moral disengagement or control aversion.

The individual's utility is thus

(3) $$u=\varphi A - g(a) + a\left(s+\lambda_0\left(1+1\{s>0\}\lambda_c + s\lambda_m\right)\right)$$

and the individual's utility maximizing contribution (a^*) equates the marginal cost of contributing to the marginal benefits (the returns from the public good plus the subsidy plus the effect on the individual's values), or:

(4) $$g'(a^*)=\varphi+s+\lambda_0\left(1+1\{s>0\}\lambda_c+s\lambda_m\right)$$

We assume that in the absence of a subsidy the contributions of the citizens to the public good given by (4) are inefficient in the sense that there exists a mutual increase in contributions that would make all citizens better off. The causal structure of the model is illustrated in Figure 3.

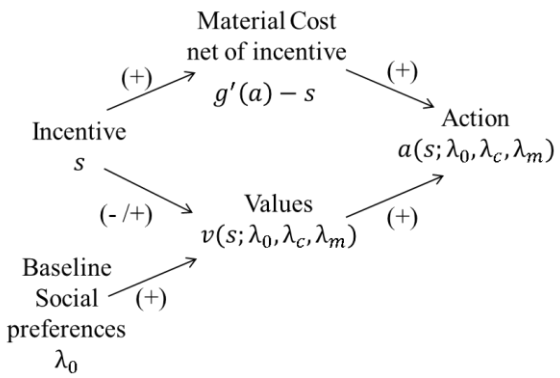


Figure 3. Crowding effects of incentives for an individual with state-dependent preferences. Baseline social preferences are the individual's non-material motivations to contribute to the public good in the absence of an incentive. Incentives reduce the net cost of contributing to the public good; but unless $\lambda_c = 0 = \lambda_m$ (separability) or $\lambda_0 = 0$ (no social preferences to crowd out) they also affect the motivational salience of the individual's social preferences.

The introduction of a subsidy increases contributions by raising the marginal benefits of contributing, the right hand side of (4) which we denote, θ. Considering the case in which there initially is no incentive, the effect of an incentive on the net benefits of contributing (expressed in discrete terms so as to be able to account for the discontinuity in the value function at $s=0$) is

(5) $$\frac{\Delta\theta}{\Delta s}=1+\lambda_0\left(\frac{\lambda_c}{\Delta s}+\lambda_m\right)$$

and is composed (as expected) of a direct effect (that is, 1, the top arrows in Figure 1),and the indirect state-dependent effect which will be negative in the case of crowding out (λ_c or λ_m negative), and larger in absolute value the greater are the baseline values of the individual (λ_0). We likewise see that

(6) $$\frac{\Delta\theta}{\Delta\lambda_0}=1+1\{s>0\}\lambda_c+s\lambda_m$$

which, in the case of crowding out, is declining in s.

Equations (5) and (6) make it clear that when λ_c and λ_m are negative, incentives and baseline values are substitutes: the effect of each on the marginal benefits of contributing varies inversely with the level of the other. The fact (from equation 5) that the crowding effect is

larger for those with greater baseline social preferences makes sense and is consistent with experiments that have identified the strength of individuals' social preferences independently of incentives and found that crowding out effects are larger for those with greater baseline values (Bohnet and Baytelman (2007), Kessler (2008), Carpenter and Myers (2010)). This substitutability between incentives and baseline values will be important when we address questions of public policy in the penultimate section.

Using (5) we say that a particular change in incentives Δs has crowded out social preferences if $\Delta\theta/\Delta s < 1$, that is, if the total effect of the incentive is less than the direct effect, and conversely for the case of crowding in. Crowding will not occur if λ_c and λ_m or λ_0 are zero (that is, if social preferences are not state-dependent, or they are absent). What we term strong crowding out holds if $\Delta\theta/\Delta s < 0$, which from the right hand side of (5), we see may occur if categorical crowding out is large relative to the size of the subsidy, or its marginal effect ($\lambda_0\lambda_m$) is small, or if the marginal effect is negative.

The two forms of non-separability are illustrated by crowding out in Figure 4. Crowding in, which we do not show, would either shift the "separability" function upwards – categorical crowding in -- or increase its slope – marginal crowding in. Because the functions in Figure 4 represent the citizens' best responses to the planner's choice of an incentive and thus constitute one of the constraints making up the planner's optimizing problem, we call these functions the planner's implementation technology.

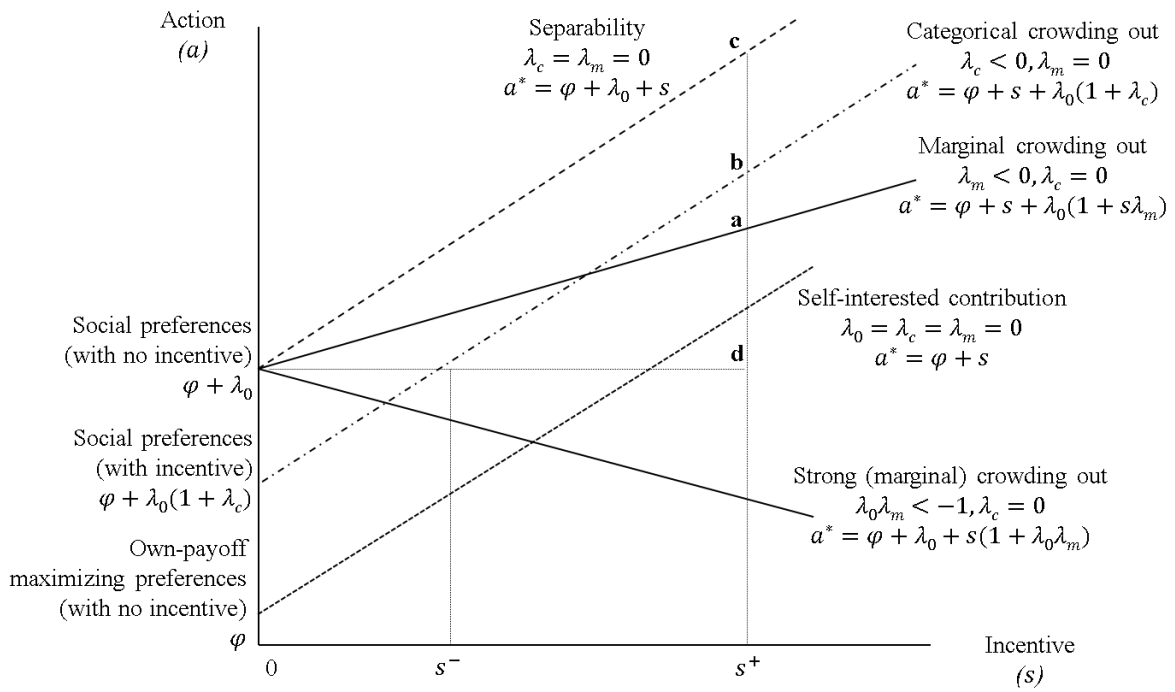


Figure 4. The sophisticated planner's implementation technology: **Citizen's contribution to the public good with state-dependent social preferences.** Under separability (top dashed line) incentives and social preferences are additive. Under strong marginal crowding out the use of the incentive is counterproductive (i.e. reduces contributions). Under categorical crowding out (dot-dashed line), incentives are also counterproductive for sufficiently small $s < s^-$.

An experiment allows an estimate of both categorical and marginal crowding out. Bernd Irlenbusch and Gabriele Ruchala implemented a public goods experiment in which the 192 German student subjects faced three conditions: no incentives to contribute and a bonus, given to the highest contributing individual, that was either high or low (Irlenbusch and

Ruchala (2008), details are in table 3 in section 5 and results are shown in Figure 5). Payoffs were such that even with no incentive individuals would maximize their payoffs by contributing 25 units. In the no-incentive case contributions averaged 37 units, or 48 percent above what would have occurred if the participants had been motivated only by the material rewards of the game. Contributions in the low-bonus case were not significantly different from the no-bonus treatment. In the high-bonus case, significantly higher contributions occurred, but the amount contributed (53 units) barely (and insignificantly) exceeded that predicted for self-interested subjects (50 units).

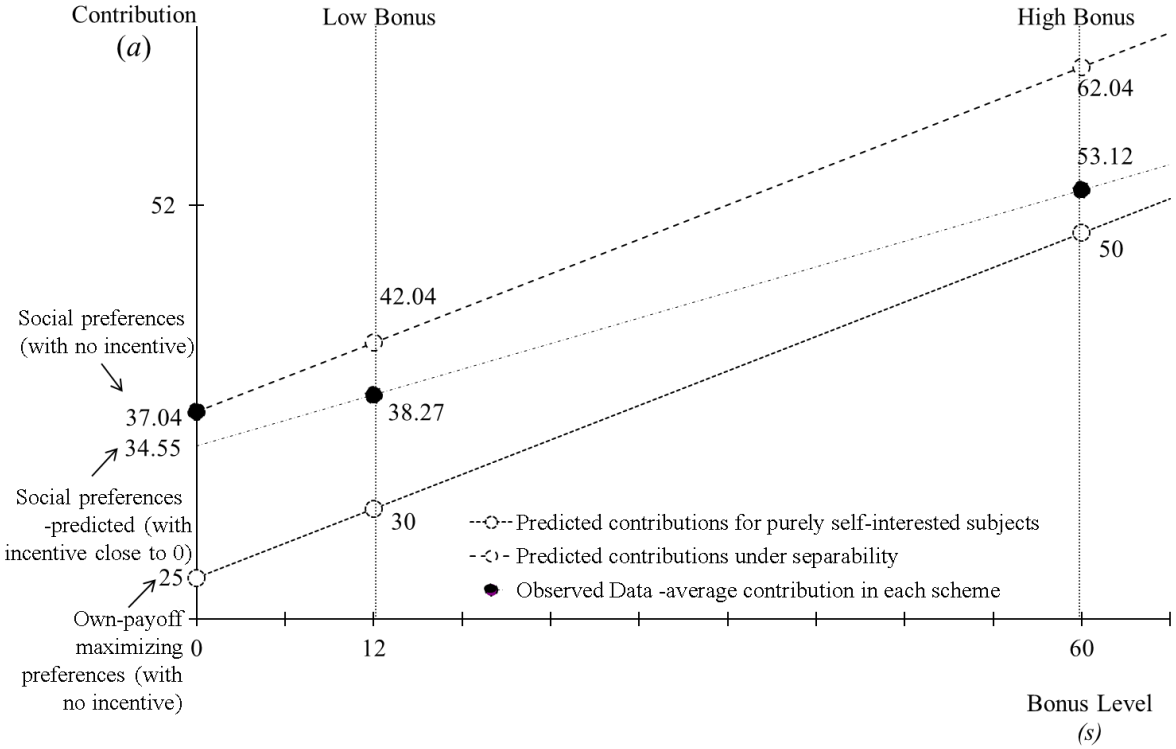


Figure 5. Categorical and marginal crowding out. Source: calculated from Irlenbusch and Ruchala (2008). See text. The experimental design is an adapted Public Goods Game comparing two team-based compensation schemes without and with a relative bonus.

In Figure 5 we use the observed behavior in the high and low bonus case along with the assumption that marginal crowding affects the slope of the citizens' best response function by a given amount (so that the function remains linear as in Figure 4) to estimate the marginal effect of the bonus. We find that a unit increase in the bonus is associated with a 0.31 increase in contributions. This contrasts with the marginal effect of 0.42 that would have occurred under separability, that is, had subjects without social preferences simply best responded to the incentive. Crowding out thus affected a 26 percent reduction in the marginal effect of the incentive. The estimated response to the incentive also gives us the level of categorical crowding out, namely the difference between the observed contributions (37.04) in the absence of any incentive and the predicted contributions had an arbitrarily small incentive been in effect (the vertical intercept of the observed line in Figure 5) or 34.55. The incentive thus categorically crowded out 21 percent of the effect of social preferences (measured by the excess in contribution levels above Nash equilibrium for self-interested subjects, 12.04).

Categorical crowding out is also evident in other experiments. In one, reported willingness to help a stranger load a sofa into a van was much lower under a small money incentive than with no incentive at all, yet a moderate incentive increased the willingness to help over the no incentive condition (Heyman and Ariely (2004). Using these data as we did in the Irlenbusch and Ruchala study, we estimate that the mere presence of the incentive reduced the willingness to help by 27 percent (compared to the no incentive condition).

Another experiment that allows us to distinguish categorical and marginal crowding was implemented by Juan Camilo Cardenas (2004), but here (as in some other experiments) we observe categorical crowding in. Cardenas implemented an experimental Common Pool Resource Game very similar in structure to the kind of real world commons problem faced by his subjects – rural Colombian eco-system users. In the absence of any explicit incentives, the villagers on average extracted 44 percent less of the experimental "resource" than would have maximized their individual payoffs, providing evidence of a significant willingness to sacrifice individual gain so as to protect the resource and raise group-average payoffs. When they were liable to pay a small fine (imposed by the experimenter) if they over-extracted the resource, as expected, they extracted even less than without the fine, showing that the fine had the intended effect.

The fact that the average extraction under the small fine treatment was 55 percent less than the Nash equilibrium for self-interested subjects (when account is taken of the fine) suggests that the fine had increased the salience of the villagers' social preferences (by 25 percent, if the 44% deviation from the self-interested Nash behavior is taken as the measure of social preferences). Interestingly, raising the fine from a low to a high level had virtually no effect. Variations in the fine thus did not work as an incentive, but rather (in Cardenas' view) the very presence of the fine (high or low) was a signal, one that alerted subjects to the public good nature of the interaction. We will present other examples of fines as signals (section 3) and crowding in (section 9). These cases hold important lessons for why incentives sometimes are counterproductive and how well-designed policies can make incentives and social preferences complements rather than substitutes.

Unfortunately, unlike the Irlenbusch and Ruchala and Cardenas studies, many experiments do not establish the response to incentives that would be observed under separability, so it is impossible to determine if incentives are "under-performing." A common misinterpretation of experimental results is to infer from the observation that an incentive has an effect in the intended direction that crowding out has not occurred (Rigdon (2009)). But observing a positive incentive effect in an experiment does not preclude crowding out. It is clear from Figure 4 (or equation 5 and the definition of crowding out) that a positive incentive effect may occur in the presence of marginal crowding out (as long as it is not "strong") and in the presence of categorical crowding out (as long as the incentive is sufficiently large.) For example, consider some substantial incentive indicated by s^+ in the figure 4. Under both marginal and categorical crowding out, the action taken (points **a** and **b** respectively) is greater than in the absence of the incentive (**d**), so the incentive "worked": it affected the action in the intended direction. But the diagnostic for the presence of crowding is a comparison of these two action levels with the level that would have occurred under separability, namely point **c**, and this comparison makes it clear that crowding out occurred.

4. Incentives alter cultural learning: a model of endogenous preferences.

A quite different mechanism by which crowding might occur has also been studied, one in which preferences are endogenous so that one or more of the parameters of the individual's value function -- λ_0, λ_c and λ_m-- are altered by incentives (Bar-Gill and Fershtman (2005), Hwang and Bowles (2011b)). Hwang and Bowles present a model of cultural evolution in which the presence or level of incentives affects the process by which preferences are acquired or abandoned, so that a population's equilibrium distribution of preferences depends on incentives. By equilibrium preferences they mean a configuration of incentives and preferences such that the latter are stationary given the process of preference-updating.

In the Hwang and Bowles model preferences are endogenous because i) schools, families, religious organizations and other societal institutions seek to promote civic minded values and ii) individuals periodically alter their preferences in response to their own recent experiences. Their model of endogenous preferences is based on two empirical regularities. The first is the powerful effect of mere exposure on preferences, documented by the social psychologist Zajonc (1968) and in subsequent works (Birch and Marlin (1982), Murphy and Zajonc (1993), Murphy, Monahan and Zajonc (1995)). The exposure effect is one of the reasons that cultural transmission may have what is termed a conformist component that favors the numerous over the rare, independently of their economic success (See Boyd and Richerson (1985): 223ff, Ross and Nisbett (1991):30ff, Bowles (1998) and the works cited there.) Following Robert Boyd and Peter Richerson, Hwang and Bowles assume a degree of conformist cultural transmission, so that the likelihood that an individual will adopt a particular preference varies not only with relative payoffs associated with the behaviors motivated by the preference but also with the prevalence of individuals with that preference in the population.

The second empirical regularity captured in their model of individual updating is that the presence and extent of incentives to contribute to a public project (or to engage in similar activities that benefit others) make the action (contribution) a less convincing signal of an individual's social preferences, resulting in observers interpreting some generous acts as merely self-interested. This is the key mechanism in the model of Roland Benabou and Jean Tirole showing how incentives may crowd out pro-social behavior (Benabou and Tirole (2006)). Similarly, in his “Generous actors, selfish actions” paper, and in his subsequent work with Dufwenberg, Heidhues, Kirchsteiger, et al. (2011), Sobel (2009) and his co authors provide not very restrictive conditions on individual utility functions such that “agents who care directly about the welfare and opportunities of others cannot be distinguished from selfish agents in market settings” (p.19). The reason is that for a class of utility functions admitting such other regarding preferences as inequality aversion, (paraphrasing the main theorem in their 2011 paper, p. 6) the “set of Walrasian equilibria of an economy [with other regarding preferences] coincides with the set of Walrasian equilibria of its corresponding ... economy [in which] agents care only about their own direct consumption.” Thus the use of market-like incentives may make it impossible to infer generous or fair-minded behaviors from the observed actions of ones fellow citizens.

There are two reasons why the presence of an incentive may lead people to mistake a generous act – helping another at a cost to oneself -- for a self-interested one. The first is that the incentive provides a competing explanation of the generous act: “he did it for the money”.

The second is that incentives often induce individuals to shift from an ethical to a payoff maximizing frame (even relocating the neural activity to different regions of the brain); and knowing this, the presence of an incentive for an individual to help another may suggest to an observer that the action was self-interested (Gneezy and Rustichini (2000b), Heyman and Ariely (2004), Irlenbusch and Sliwka (2005), Li, Xiao, Houser, et al. (2009)). The first "he did it for the money" reason depends on the magnitude of the incentive because in order to provide a convincing self-interested interpretation for the helping act the subsidy would have to exceed the cost of helping. The second reason --"when incentives are in force, everyone maximizes their payoffs" -- is categorical; it is simply the presence of the incentive that matters. Of course, an observer could make the opposite mistake, inferring that the generous act that was motivated entirely by an incentive, was done for ethical rather than payoff maximizing reasons. In the model that follows the incentive is assumed on balance to degrade helping as the signal of a generous individual's type rather than motivating self-interested individuals to act in ways that are mistakenly taken as signals of a generous type.

Taken together, these two assumptions imply that the extensive use of incentives may reduce the perceived population frequency of individuals with social preferences, leading (via the conformist learning effect) to an evolutionary disadvantage of generosity over self-interest in the preference-updating process. To show this Hwang and Bowles (2011b) adapt the model of endogenous preferences in Bowles (1998) and (2004) to study the effects of incentives on the preference-updating process. In terms of the state-dependent model of the previous section, they study the effect of incentives on the equilibrium fraction of the population for whom λ_0 is positive and sufficient to motivate contribution to a public good. The causal structure of their model is shown in figure 6.

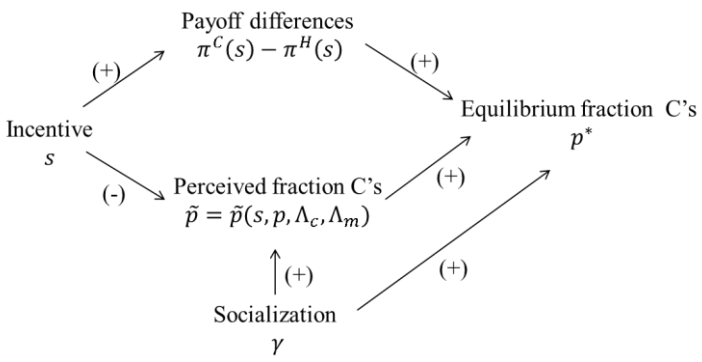


Figure 6. Crowding effects of incentives for a population with endogenous preferences. Incentives raise the relative payoff of those contributing to the public good, supporting a larger fraction of civic-minded citizens; but unless Λ_c=0=Λ_m (separability) or $\underline{p} = 0$ in which case $p^* = 0$ (no civic minded citizens in the absence of a subsidy) a subsidy also alters the preference updating process by reducing the perceived fraction of civic-minded citizens $\tilde{p}$.

Suppose there are two types: a Civic gives to the public good at a personal cost equal to $g > 0$ that may be partially offset by a subsidy s, while *Homo economicus* does not contribute and receives no subsidy. Both types update their traits by myopic best response, observing the material payoffs and public goods contribution of a sample of the population (they do not observe the utility of others) and a signal $\tilde{p}$ (possibly inaccurate when the planner implements a subsidy, $s > 0$) of p, the true frequency of the Civics in the population, p,so $\tilde{p} = \tilde{p}(s, p)$ which is decreasing in s. To capture the fact that the effect of the incentive on citizen's perception of the fraction of their fellow citizens who are Civics may depend on the mere presence of the incentive or on its extent, (or both), let

$$\tilde{p} = p\left(1 + \mathbf{1}\{s > 0\}\Lambda_c + s\Lambda_m\right) \tag{7}$$

where as before the indicator $\mathbf{1}\{s > 0\} = 1$ if $s > 0$ and zero otherwise and $\Lambda_c \leq 0$ measures the categorical effect of the presence of an incentive on one's inference about another individual's type based on observing his or her contribution to the public good and $\Lambda_m \leq 0$

measures the marginal effect of the level of an incentive on one's inference. Note that when $p = 0$ or $s = 0$, $\tilde{p} = p$ so in the absence of the subsidy or when Civics are absent, the citizen's perception of the fraction of the population who are Civics is accurate.

The incentive has two offsetting effects on the distribution of types in the population, one intended and the other not: it raises the relative payoffs of the Civics, but it also reduces their apparent prevalence in the population. To see how this affects the equilibrium distribution of types in the population, suppose that individuals live forever but they periodically may switch their type. Denote the cultural fitness of trait i as r^i ($i = C, H$ for Civic and *Homo economicus*) defined as the expected number of replicas that each individual bearing the trait will leave in the subsequent period. (If person k switches to j's type and j does not switch then k has left no replica and j has two replicas.) To capture the effect of socialization institutions on the evolution of preferences in this population the authors suppose that in any period some fraction γ of the H-types will be converted to a C-type. (Because it plays little role in what follows, Hwang and Bowles do not model the manner in which socialization institutions accomplish this, other than to assume that the process is not affected by the level of incentives). They define $\alpha \in (0,1]$ as the relative weight of conformism rather than payoffs in the updating process, β as the weight of payoff differences relative to the socialization effect and π_C, π_H as the expected payoffs of the two types. Then, taking account of the common influences on the number of replicas, r_0, the cultural fitness of the two traits can be written:

(8) $$r^C = r_0 + a\left[\left(\tilde{p} - \frac{1}{2}\right) + (1-a)\left[b\left(p_C - p_H\right) + g\frac{1-p}{p}\right]\right]$$

(9) $$r^H = r_0 + a\left[\left(\frac{1}{2} - \tilde{p}\right) + (1-a)\left[b\left(p_H - p_C\right) + g\right]\right]$$

The second term in both equations is the conformism effect, and it favors the Civics if it is perceived that they constitute more than half of the population. The third term is the net effect of socialization and payoff based updating. The socialization effect in equation (8) (the second term in the square brackets) is derived as follows: noting that population size is normalized to unity, each of the $1 - p$ H types in the population has a γ probability of converting to C (shown in (9)) and thus appearing as $\gamma(1 - p)$ replicas assigned to the p Cs in the population. The final expression in (8) is thus the per C share of these socialized former H's.

From these cultural fitness equations one readily derives the familiar replicator equation for the movement of p over time:

(10) $$dp/dt = p\left(1 - p\right)\left(r^C - r^H\right)$$

Taking account of the costs of contributing to the public good and the subsidy, and noting that the payoff difference between the types $\pi_H - \pi_C$ is just $g - s$, the resulting stationary condition for an interior value of p (namely $r^C - r^H = 0$, so that $dp/dt = 0$) is

(11) $$\left(\tilde{p} - \frac{1}{2}\right)\frac{a}{1-a} = b\left(g - s\right) - \frac{g}{2p}$$

which requires that the conformist effect favoring the more common trait (the left hand side) offset the net effect of that trait's payoff disadvantages and the societal level socialization effects (the right hand side). Values of p satisfying (11) are termed the population's equilibrium preferences and denoted as $p^*(s)$.

Figure 7 illustrates the cultural equilibrium condition (11). The solid lines show the two sides of equation (11) – the conformist effect and the payoff plus socialization effects -- and their intersection, satisfying equation (11) when $s = 0$ and giving $p^*(0)$ that is, the equilibrium distribution of preferences in the absence of incentives. The dotted lines show the effect of the implementation of a subsidy. The intended effect is to reduce the payoff advantage of the H types (they do not receive the subsidy) shifting downward the payoff cum socialization function. A naive social planner, unaware of the conformist effect would thus expect the introduction of the incentive to increase the fraction of C's in the population to $p^N(s)$.

But the unintended effect of the subsidy is to reduce the perceived fraction of the population who are C's and thereby to diminish the conformist advantage of the C's. The downward shift in the conformist effect function thus partially offsets the payoff effect, with the resulting stationary distribution equal to $p^*(s)$. In the case of strong crowding out (not shown), the second effect would more than offset the first, resulting in a $p^*(s) < p^*(0)$.

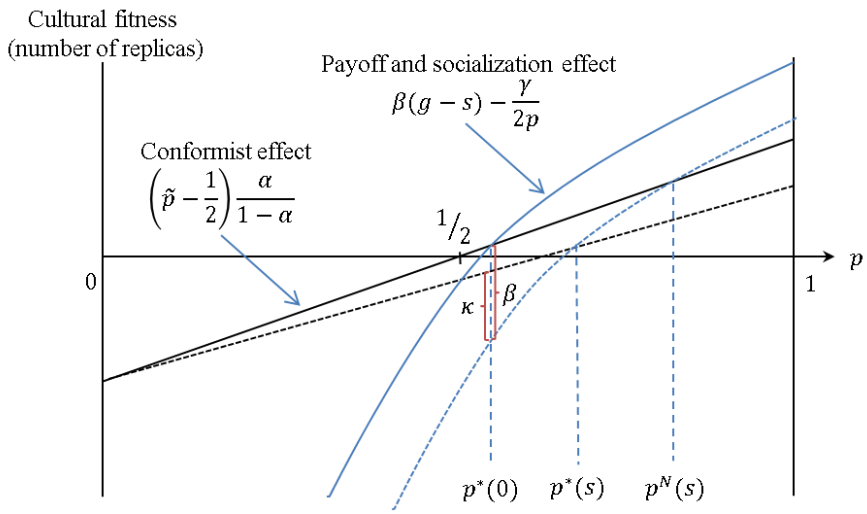


Figure 7. Incentives and equilibrium preferences. The figure (solid lines) shows the determination of fraction of citizens with social preferences in a cultural equilibrium under the influence of payoff-based and conformist updating, namely p^* when $s = 0$.The subsidy (dotted lines) reduces the payoff difference between *Homo economicus* and Civics, and in the absence of the effect on the perceived frequency of conformism in the population, the fraction of C's in the population would increase from $p^*(0)$ to $p^N(s)$. However, the associated reduction in the conformism effect partially offsets this. The resulting equilibrium outcome is $p^*(s)$. The distance indicated by β is the effect anticipated by the naive planner who assumes separability, while κ is the true effect taking nonseparability into account. Source: Hwang and Bowles (2011b)

The source of the non-separability between socializations and incentives is clear if we return to equation (11) and consider the effect of an increase in s on the cultural fitness of the C types relative to the H types, evaluated at the status quo distribution of types in the population. In the initial of any subsidy, is just the vertical distance at $p^*(0)$ between the two functions that have been displaced by the introduction of the incentive (the dashed lines). Because this effect is the cultural fitness advantage of the C-types following the introduction of the incentive, when $p = p^*(0)$, Hwang and Bowles term it the evolutionary impact of the incentive, denoted by κ (analogous to θ in the state-dependent model of the previous section, κ is the difference between the two sides of (11)). The direct effect of incentives on κ is just β, but as is clear from the following expression, there also is an indirect effect:

$$\left.\frac{\Delta\kappa}{\Delta s}\right|_{p^*} = \beta + \frac{\alpha}{1-\alpha}\, p^*(\gamma, s)\left[\Lambda_m + \frac{\Lambda_c}{\Delta s}\right] \tag{12}$$

where the left hand side means the change in the equilibrium condition associated with the change in s, for the given level of p, namely p^*. (As in the model of state-dependent preferences, we consider discrete changes here rather than simply differentiating (11), in this case because of the discontinuity of $\tilde{p}$ at $s = 0$ in the presence of categorical crowding.)

The indirect effect will be negative in the case of crowding out, so the total effect of the incentive is less than the direct effect. The absolute size of the indirect effect (the second term on the right of (12)) is (as expected) increasing in the extent of conformism in updating and in the (absolute magnitude of the) crowding parameters. Importantly, the negative indirect crowding out effect will be larger in absolute value, the greater is p^*. Because p^* varies positively with the socialization effect (γ), the total effect of the incentive is less, the more effective is a society's socialization institutions. The crowding effect will absent (separability will hold) if $\Lambda_m = \Lambda_c = 0$ in which case $\tilde{p} = p$, in which case there are no misperceptions of the fraction of C's in the population, or $\gamma = 0$ so there are no C's to misperceive as self-interested, or $\alpha = 0$ in which case there is no conformism in updating so the mis-perceptions induced by the incentives have no effect.

We also have that the evolutionary impact of socialization institutions is

$$\left.\frac{\Delta k}{\Delta g}\right|_{p^*} = \frac{1}{2p^*(g,s)} \tag{13}$$

which diminishes with greater use of incentives because (in the absence of strong crowding out) incentives raise p^*.

Thus where crowding out occurs incentives and socialization institutions are substitutes in the sense that the marginal effect of one on the evolutionary advantages of the civic minded types diminishes with the level of the other. We will return to the property of incentives and socialization as substitutes and the possibility of making them complements when we consider the policy implications of these models and the data to follow. A summary of the two sources of non-separability – state-dependence and endogeneity of preferences – and the mechanisms involved is provided in Table 2.

Preferences	**State-dependent**	**Endogenous**
Exogenous determinant of social preferences	Individual baseline values λ_0	Population level socialization effect γ
Crowding mechanism	Salience of values $v = \lambda_0\left(1 + \mathbf{1}\{s > 0\}\lambda_c + s\lambda_m\right)$	Perceived fraction C's $\tilde{p} = p\left(1 + \mathbf{1}\{s > 0\}\Lambda_c + s\Lambda_m\right)$
Intended target of the incentive	Individual best response $a^*\left(s, g(a), \lambda_0, \lambda_c, \lambda_m\right)$	Fraction of Cs in population $p^*\left(s, g, g, \Lambda_c, \Lambda_m\right)$
Separability	$\lambda_0\left(\frac{\partial \lambda_c}{\partial s} + s\lambda_m\right) = 0$	$\frac{a}{1-a} p^*(g,s)\left(\Lambda_m + \frac{\Lambda_c}{\partial s}\right) = 0$
Sufficient Conditions: separability	$\lambda_0 = 0$ or $\lambda_c = \lambda_m = 0$	$\gamma = 0, \alpha = 0$ or $\Lambda_c = \Lambda_m = 0$
Necessary Conditions: crowding out (in)	$\lambda_0 > 0$, λ_c or $\lambda_m < (>) 0$	$\gamma > 0$, $\alpha > 0$ Λ_c or $\Lambda_m < (>) 0$

Table 2 Separability and crowding when social preferences are state-dependent or endogenous. In both models the citizens may bear a cost ($g(a)$ or g) in order to contribute to a public good where a subsidy, s, may partially offset the cost. In the endogenous preference model those who contribute are C's. Additional notation: λ_m , λ_c and Λ_m ,Λ_c are the marginal and categorical crowding parameters (in the state-dependent and endogenous cases, respectively) and α is the relative importance of conformism in the endogenous preferences model.

The design of effective incentives in cases where separability may not hold requires a better understanding of the cognitive or affective effects of incentives that explain the categorical and marginal crowding out effects observed in experiments. We turn in the next three sections to the mechanisms that make preferences incentive-state-dependent, resulting in crowding out effects before considering (in section 8) the evidence for the adverse effects of incentives on preference-updating. (We consider crowding in --the case where incentives and social preferences are complements-- in section 9.)

5. Bad news: Incentives provide information about the principal

Incentives are implemented for a purpose, and because the purpose is often evident to the target of the incentives, the target may also infer information about the person who designed the incentive, about his or her beliefs concerning the target, and the nature of the task to be done (Benabou and Tirole (2003), Fehr and Rockenbach (2003)). We will illustrate this incentives-as-information-about-the-incentive designer effect by the negative response to fines imposed by experimental 'investors' and 'trustees' in the Trust Game, a principal-agent experiment implemented by Ernst Fehr and Bettina Rockenbach.

German students in the role of "investor" were given the opportunity to transfer some amount to the other player, called the "trustee". This amount was then tripled by the experimenter. The trustee, knowing the investor's choice, could in turn "back-transfer" some (or all, or none) of this tripled amount, returning a benefit to the investor (Fehr and Rockenbach (2003)). When the investor transferred money to the trustee, he or she also specified a desired level of the back-transfer. The experimenters implemented an incentive condition in which the investor had the option of declaring that he would impose a fine if the trustee's back-transfer were less than the desired amount. The investor could also decline the use of the fine, the choice of using or declining the fine option being known to the trustee and taken prior to the trustee's decision. There was also a "trust" condition in which no such incentives were available to the investor.

Trustees reciprocated generous initial transfers by investors with greater back-transfers. But the use of the fine reduced return transfers conditional on the investor's transfer, while renouncing the use of the fine when it was available to the investor increased back-transfers. Only one-third of the investors renounced the fine when it was available; their payoffs were 50 percent greater than the investors who threatened use of the fines.
The proximate causes of the negative impact of incentives in this case are suggested by evidence on the neural responses of the trustees in another Trust Game experiment (Li, et al. (2009)) As in the Fehr and Rockenbach experiment, the investor's threat of sanctions negatively affected back-transfers by trustees. To identify the proximate causes of this result, Jian Li and his co-authors used functional magnetic resonance imaging (fMRI) to compare the activation of distinct brain regions of trustees when faced with an investor who had threatened to sanction the trustee for insufficient back-transfers and an investor who had not threatened a sanction. Sanction threats de-activated the Ventromedial Prefrontal Cortex (VMPFC), a brain area whose activation was greater in trustees who made larger back-transfers, as well as other brain areas thought to be involved in the processing of social rewards. The threat activated the parietal cortex, an area thought to be associated with cost-benefit analysis and other self-interested optimizing processes. The interpretation by Li and his coauthors is that the sanctions induced a "perception shift" favoring a more self-interested response.

The signaling interpretation of counter-productive incentives in the Trust Game suggested by Fehr and Rockenbach is that in the trust condition, or when the fine was renounced by the investor, a large initial transfer signaled that the investor trusted the trustee. The positive response to the investor's renunciation of the fine option is a categorical effect, analogous to the negative categorical effect of the use of incentives in the Irlenbusch and Ruchala experiment described above. The threat of the fine, however, conveyed a different message and diminished the trustee's reciprocity.

Similar cases of crowding out due to the "bad news" conveyed by the incentive are at work in experiments among student subject pools in Switzerland, U.S., Italy, France and Costa Rica (as well as Germany) and in a diverse set of games including Gift Exchange, Public Goods, and a charity giving setting similar to a Dictator Game. Costa Rican businessmen also responded negatively to the bad news that incentives conveyed. Table 3 summarizes experiments in which this incentives-as-signals effect appears to have been at work (in some cases along with other mechanisms, to which we now turn [16, 17, 19, 24, 28]). Crowding out as the result of the "bad news" mechanism may be prevalent in Principal Agent settings and can be averted where the principle has a means of signaling trust or fairness (experiments [1-3]). Not surprisingly crowding out affects individuals who are intrinsically motivated or fair-minded [5-6]; for own payoff maximizers, it appears there is nothing to crowd out.

6. Moral disengagement: Incentives may suggest permissible behavior

In most situations people look for clues of appropriate behavior, and incentives often provide them. In Table 4 we survey experiments in which this framing effect appears to have been at work. These experiments differ from those in Table 3, in which incentives were deployed by experimental subjects in the role of a principal interacting with an agent. Here incentives are implemented exogenously, that is by the experimenter, so that they provide no information about the intentions or beliefs of other experimental subjects. As can be seen from the table, incentives appear to affect moral disengagement not only among students but also (as we have seen) among poor Colombian villagers [13, 14] and top U.S. CEOs [16]. Moral disengagement was evident in the Ultimatum Game and the Common Pool Resource Game [11-14; 20] as well as in the games for which the bad news mechanism was at work [1,3,6,7,9,10]. In addition, this mechanism may be clearly recognized in settings of 1-player games (i.e. Dictator Game or a performance Task) [30, 32].

Hoffman and her co authors (1994) illustrated the framing power of names: generosity and fair-minded behavior were diminished by simply re-labeling an Ultimatum Game the "Exchange Game" and re-labeling proposers and responders "sellers" and "buyers" (Hoffman, McCabe, Shachat, et al. (1994)). The power of names has been confirmed in many (but not all) experiments since then (Zhong, Loewenstein and Murnighan (2007)) but in some cases (Ellingsen, Johannesson, Möllerström, et al. (2011)) the framing effect appears to have altered subjects beliefs about the actions of others rather than their preferences.

But literally naming the game is not necessary for framing effects to occur. Incentives alone may provide powerful frames for the decision maker. A year before the first reality TV Survivor show, Schotter and his coauthors found that market-like competition for "survival" among subjects reduced their concern for fairness in an Ultimatum Game experiment (Schotter, Weiss and Zapater (1996)). In this game Player 1 is given an endowment and asked

to propose a division of it with Player 2. Player 2, knowing the size of the endowment, decides whether to accept or reject the division. If Player 2 accepts, then the proposed division is implemented. If Player 2 rejects, both players receive zero. As is commonly observed in the Ultimatum Game, Player 1 made quite generous offers and low offers were frequently rejected. But when the experimenters told the subjects that those with lower earnings would be excluded from a second round of the game, Player 1 subjects offered less generous amounts to Player 2, and Player 2 accepted lower offers. The authors' interpretation was that: "...the competition inherent in markets...offers justifications for actions that, in isolation, would be unjustifiable." (p.38).

While plausible, direct evidence for this "moral disengagement" explanation is lacking because the social preferences that apparently accounted for fair behavior in the non-survival condition of the experiment were not measured. There are cases, however, in which the reduction in the salience of ethical reasoning induced by the presence of incentives can be detected.

A large team of anthropologists and economists implemented both Dictator and Third Party Punishment Games in 15 societies ranging from Amazonian, Arctic and African hunter gatherers to manufacturing workers in Accra, Ghana and US undergraduates (Barr, Wallace, Ensminger, et al. (2009), Henrich, Ensminger, McElreath, et al. (2010)). In the Dictator Game an experimental subject is assigned a sum of money and asked to allocate some all or none of it to a passive recipient. The Third Party Punishment Game is a Dictator Game with an active onlooker (the third party) who observes the dictator's allocation. If the third party deems the dictator's allocation worthy of punishment he or she may then pay to impose a monetary fine on the dictator. One would expect that in the presence of a third party, the dictators would adjust their allocations upwards (compared to the two party standard Dictator Game) and thus avoid being fined. But this was not the case; fining was common; it occurred in 30% of the interactions across the study sites.

Surprisingly, in only two of the 15 populations were the offers significantly higher in the Third Party Punishment Game than in the Dictator Game, and in four of the populations the allocations were significantly (and in some cases substantially) lower. In Accra, for example, where 41 percent of the dictator's allocations resulted in fines by the third party, the allocations were 30 per cent lower in the Third Party Punishment Game than in the Dictator Game. The incentives provided by the fine did not induce higher allocations, but rather had the opposite effect. (For two groups there was a significant positive effect of the fine option indicating that in these two cases the incentive had some effect; but as we have seen does not preclude crowding out.)

Crowding out of ethical motives is suggested by the fact that the dictator's adherence to one of the world's religion (Islam or Christianity, including Russian Orthodoxy) raised allocations in the Dictator Game by 23 percent (compared to those unaffiliated with a world religion.). But in the Third Party Punishment Game, the estimated "religion effect" was reduced to just 7 percent of its value in the Dictator Game and it was not significantly different from zero. The presence of the incentive based on the fine appears to have defined the setting as one in which the moral teachings of these religions were not relevant. Consistent with a crowding out interpretation of these results, the negative effect on the dictator's allocations of his or her economic need (number of children, conditional on a given level of income and wealth) was substantial (and statistically significant) in the Third Party Punishment Game, but in the Dictator Game this "economic need effect" was an order of

magnitude smaller and not significantly different from zero.

In the Accra sample (Barr (2004)) the dictator's allocation co-varied significantly with the frequency of attendance at church or mosque in the standard two party Dictator Game; but this large "religion effect" vanished in the Third Party Punishment Game. The incentives implicit in the Third Party Punishment Game appear to have substituted economic motivations for moral concerns. These experiments are also consistent with our model of state-dependent preferences, in which crowding out operates via an effect of incentives on the behavior of those with pre-existing social preferences.

7. Control aversion: Incentives may compromise intrinsic motives and self-determination

Recent experiments by economists surveyed in Table 5 as well as non-experimental studies in economics (surveyed in Frey and Jegen (2001)) provide evidence for a third reason why social preferences may be state-dependent in ways leading to crowding out. Table 5 does not include the original "over-justification" experiments done by psychologists (referred to in the introduction). Unlike the experiments by psychologists where incentives are typically implemented by the experimenter, economists often model strategic interactions in which the same apparently control averse reaction occurs, so these experiments could also fall under the "bad news" about the principal rubric presented in Table 3 [6,10]. Moreover, framing effects may result in moral disengagement in some of these experiments [21, 26, 28, 31]. Crowding out effects of intrinsic motivation may be recognized in Ultimatum games [11, 12] and games where the experimenter is the principal [18, 28, 31]. We think it is likely that in these and other cases more than one mechanism is at work.

Armin Falk and Michael Kosfeld used a principal-agent game to explore the idea that 'control aversion' based on the self-determination motive may be a reason why incentives sometimes degrade performance (Falk and Kosfeld (2006)). Experimental agents in a role similar to an employee chose a level of 'production' that was costly to them and beneficial to the principal (the employer). The agent's choice effectively determined the distribution of gains between the two, with the agent's maximum payoff occurring if he produced nothing. Before the agent's decision, the principal could elect to leave the choice of the level of production completely to the agent's discretion, or impose a lower bound on the agent's production (three bounds were varied by the experimenter across treatments, the principal's choice was simply whether or not to impose it.) The principal could infer that a self-interested agent would perform at the lower bound or, in the absence of the bound, at zero, and thus imposition of the bound would maximize the principal's payoffs.

But in the experiment agents provided a lower level of production when the principal imposed the bound. Apparently anticipating this response, fewer than a third of the principals opted for its imposition in the moderate or low-bound treatments. This minority of "untrusting" principals earned on average half of the profits of those who did not seek to control the agents' choice in the low-bound treatment, and a third less in the intermediate bound condition.

Control aversion and the desire for self-determination are not the only effects of the principal's seeking to bind the agent. As anticipated by our discussion of the information content of incentives above, the imposition of the minimum in this experiment gave the

agents remarkably accurate information about the principals' beliefs about them. In post-play interviews, most agents agreed with the statement that the imposition of the lower bound was a signal of distrust; and the principals who imposed the bound in fact had substantially lower expectations of the agents. The untrusting principals' attempts to control the agents' choices induced over half of the agents (in all three treatments) to contribute minimally, thereby affirming the principals' pessimism. Depending on the distribution of principal's priors about the agents, a population with preferences similar to these experimental subjects could support both a trusting and an untrusting (Pareto-inefficient) equilibria. Thus results in the Falk and Kosfeld experiment appear to be the result of both compromised self-determination and negative information about the incentive designer.

8. The economy produces people: Incentives alter how new preferences are learned

As in the Hwang and Bowles model introduced in section 4, incentives may also affect long-term change in motivations because they alter key aspects of how we acquire our motivations, influencing both the range of alternative preferences to which one is exposed and the economic rewards and social status of those with preferences different from one's own (Bisin and Verdier (2001), Bowles (2004), Bar-Gill and Fershtman (2005)).

Experiments of at most a few hours duration are unlikely to uncover the causal mechanisms involved in this process of durable preference change. This is because adopting new preferences is often a slow process more akin to acquiring an accent than to choosing an action in a game. As in the Hwang and Bowles model above, the The developmental processes involved typically include population-level effects such as conformism, schooling, religious instruction and other forms of socialization that are not readily captured in experiments. Acquiring new preferences (like a new accent) often takes place early in life and the learning process is strongly attenuated thereafter.

However, historical, anthropological, social psychological and other data (surveyed in Bowles (1998)) provide evidence for endogenous preferences, showing that economic structures affect parental child rearing values, personality traits rewarded by higher grades in school, and other developmental influences. Additional evidence that preferences are endogenous comes from the experimental studies of 15 small scale societies with extraordinarily varied economic structures, ranging from farming to hunting and gathering. In these studies cross subject pool comparisons showed a strong association between the nature of the diverse economic tasks required to secure a livelihood – participating in large cooperative hunting teams in contrast to solitary work in forest slash and burn horticulture, for example -- and its members' experimentally measured generosity and fair-mindedness in the Ultimatum Game (Henrich, Boyd, Bowles, et al. (2005), Henrich, et al. (2010)).

Despite the limitations of experiments for the investigation of preference change, we survey in Table 6 a number of experiments that are consistent with durable learning effects of incentives. (We have placed all of the experiments consistent with preference endogeneity in this table; of course many of them also provide evidence of the mechanisms we have identified as affecting state-dependent preferences.) We take as evidence for this the fact that the apparent effect of incentives on preferences persists even when, in later stages of an experiment, incentives are withdrawn, suggesting that the prevalence of social preferences in a population may depend on exposure to incentives in the past, as in the Hwang and Bowles

model.

An example follows. In the public goods experiment designed by Josef Falkinger, Fehr, Gächter, et al. (2000) an incentive mechanism induced subjects to contribute almost exactly the amount predicted for a self-interested individual, while in the absence of the incentive subjects contributed significantly more than would have been optimal for a self-interested individual. But, consistent with a change in preferences due to exposure to incentives, in the absence of incentives, subjects who had previously experienced the incentive system contributed 26 per cent less than those who had never experienced it.

While the cultural diversity and variety of games appearing in Table 6 are substantial, and we think the preference learning effects that we have detected in these experiments are indeed at work, we do not yet have experiments capable of testing the mechanism underlying the models of the influence of incentives on the evolution of preferences proposed by Hwang and Bowles, Bar-Gill and Fersthman, and others.

9. Crowding in

In section 2 we identified a number of cases in which crowding in may occur. For example the incentive may provide good news about the principal or it may lead to moral engagement rather than its opposite. In Table 7 we survey a number of studies that show this result. These experiments are of special interest to the social planner not only because they would ideally point the way to the design of policies which would make incentives and social preferences synergistic (that is complements) rather than substitutes, but also because it appears that crowding in occurs more often in games with more than 3 players (Public Goods [38, 41, 42, 44; 47-49] and Common Pool Resource [14, 43] games) a common characteristic of public policy settings. In the penultimate and final section we will return to these questions when we consider the policy implication of non-separability.

Synergy between incentives and social preferences may explain why fines imposed on free riders by altruistic peers in a Public Goods Game induce higher levels of contribution in subsequent rounds of play (Fehr and Gächter (2000)). Of course crowding in need not have been involved; individuals might have simply best-responded to the anticipated loss in payoffs associated with low contributions. But more than this is at work. Consistent with the interpretation that incentives imposed by peers activate shame or other social preferences, purely verbal messages of disapproval have a substantial positive effect on free riders' subsequent contributions (Barr (2001), Masclet, Noussair, Tucker, et al. (2003)). When those who have contributed more than others are punished (as sometimes occurs, Herrmann, Thoni and Gächter (2008a)), they subsequently contribute less, and costly retaliatory punishment sometimes results (Bowles and Gintis (2005), Carpenter, Bowles, Gintis, et al. (2009), Hopfensitz and Reuben (2009)). This appears to occur because the targets of the punishment feel hostility rather than shame.

There are also other mechanisms at work. The incentives and constraints typical of the rule of law and other institutional designs that limit the most extreme forms of anti-social behavior and facilitate mutually beneficial interactions on a large scale may enhance the salience of social preferences by assuring people that those who conform to moral norms will not be exploited by their self-interested fellow citizens (Bowles (2011)). This may explain the Hokkaido University subjects who cooperated more in a public goods experiment when assured that others who did not cooperate would be punished (Shinada and Yamagishi

(2007)) despite the fact that this had no effect on their own material incentives (those told this were not subject to the punishment.) They apparently wanted to be cooperative but wished even more to avoid being exploited by defectors. According to this interpretation, the fine imposed by the experimenter on any free riding liberated the individual to act pro-socially without fearing being exploited by less cooperative players. The respondents may have exhibited what Iris Bohnet and her co authors call "betrayal aversion," which was attenuated by knowing that betrayal would be punished by a third party (Bohnet, Greig, Herrmann, et al. (2008)).

Market interactions may also favor the endogenous evolution of social preferences. In two sets of experiments in small-scale societies in Africa, Asia and Latin America (Henrich, et al. (2005), Henrich, et al. (2010)), individuals from the more market-integrated societies gave more in the Ultimatum Game. The authors conjecture that this may be due to the fact that more market exposed subjects had the experience of mutually beneficial exchanges with strangers, much like in the anonymous experimental settings. A very different piece of evidence consistent with this interpretation is that subjects who were exposed to unobtrusive priming with words relating to markets and exchange prior to playing a Trust Game were more likely to trust their partner than were subjects exposed to primes unrelated to markets (Al-Ubaydli, Houser, Nye, et al. (2011)).

A distinct mechanism underlying crowding in was apparently at work in a public goods experiment by Vertova and Galbiati (2010). Consistent with the Cardenas experiment described in section 2, they found that the effect of a stated (non-binding) obligation to contribute a certain amount was greater when it was combined with a weak monetary incentive than when no incentives were offered. A stronger monetary incentive did not result in an increase in contributions. The strong monetary incentive also had no effect on behavior in the absence of the stated obligation The authors' interpretation (like that of Cardenas) is that the explicit incentives enhanced the salience of the stated obligation. In our taxonomy it is a case of categorical crowding in (See also Galbiati and Vertova (2008)).

10. The lab and the street: Can one generalize from experimental evidence?

The experimental evidence for non-separability would not be very interesting if it did not reflect real-life behavior. Testing for separability in natural settings is difficult, but generalizing directly from experiments even for phenomena much simpler than separability is a concern in any empirical study (Falk and Heckman (2009)) and is often unwarranted (Levitt and List (2007)). Consider, for example, the Dictator Game: typically more than 60 percent of the dictators allocate a positive sum to the recipient, and the average given is about a fifth of the endowment. We would be sadly mistaken if we inferred from this that 60 percent of individuals would spontaneously transfer funds to an anonymous passerby, or that the same subjects would offer a fifth of the bills in their wallet to a person who is homeless asking for help. Another example: while pro-social behavior in an experiment by Benz and Meier (2008) was correlated with non-experimental behavior, subjects who reported that they had never given to a charity allocated 65 percent of their endowment to a named charity in a lab experiment.

A possible explanation of these discrepancies between experimental and real world behavior is that most individuals are strongly influenced by the cues of appropriate behavior offered by

the situation in which an action is taken (Ross and Nisbett (1991)), and there is no reason to think that experiments are an exception to this context-dependent aspect of individual behavior. External validity concerns arise from four aspects of human behavioral experiments that do not arise in most well-designed natural science experiments. First, experimental subjects typically know they are under an unknown researcher's microscope, possibly inducing different behaviors than would occur under total anonymity or under the scrutiny of neighbors, family or workmates. Second, experimental interactions with other subjects are typically anonymous and without opportunities for ongoing face to face communication, unlike many social interactions of interest to economists and policy makers. Third, subject pools may be quite different from the real-world populations of interest, in part due to the process of recruitment and self-selection. Finally, many of the experiments that provide evidence for the salience of social preferences are deliberately structured as strategic interactions like the Ultimatum or the Public Goods Game that give scope for ethical or other-regarding behavior that may be absent in competitive markets and other important real world settings (Sobel (2010)).

It is impossible to know whether these four aspects of behavioral experiments bias experimental results in ways relevant to the question of separability. For example, the fact that in most cases subjects are paid a "show up fee" to participate in an experiment might attract the more materially oriented who may be less motivated by social preferences subject to crowding out; or knowing that the topic of the experiment was cooperation the subjects might be atypically civic minded.

We can do more than speculate about these problems. Nicole Baran and her coauthors asked if University of Chicago Graduate School of Business students who were more reciprocal in the Trust Game (those who as trustees most generously reciprocated large transfers by the investor) were also those most likely to contribute to the University upon graduation. They were (Baran, Sapienza and Zingales (2010)). Fehr and Lorenz Goette found that in a group of bicycle messenger workers in Zurich, those who exhibited loss aversion in a laboratory experiment exploring the subjects' preferences over lotteries also exhibited loss aversion when faced with real-life wage rate changes (Fehr and Goette (2007)). Dean Karlan (2005) implemented a Trust Game among Peruvians participating in a micro-credit program; those who were least trustworthy (transferred less back to the "investor") in the experiment were less likely to repay their real world loans. Alain Cohn and his co authors (Cohn, Fehr and Goette (2011)) found that reciprocators in the lab (measured by play in a sequential PD game) responded positively to a randomly awarded fixed wage increase in their work, while those who played the sequential PD in a payoff maximizing way did not respond to the wage increase.

Among the Japanese shrimp fishermen that Jeffrey Carpenter and Erika Seki studied, those who contributed more in a public goods experiment were more likely to be members of cooperatives that shared costs and catch among many boats than to fish under the usual private boat arrangements (Carpenter and Seki (2010)). A similar pattern was found among fishermen in the Brazilian north east, where some fish offshore in large crews whose success depends on cooperation and coordination, while those exploiting inland waters fish singly. The ocean fishers were significantly more generous (in Public Goods, Ultimatum and Dictator Games) than the inland fishers (Leibbrandt, Gneezy and List (2010)).

A better test of the external validity of experiments would include a behavior-based measure of how cooperative the individuals were, not simply whether they took part in a cooperation-

sensitive production process. The Brazilian fishers provide just such a test. Shrimp are caught in large plastic bucket-like contraptions; holes are cut in the bottom of the traps to allow the immature shrimp to escape, thereby preserving the stock for future catches. The fishermen thus face a real world social dilemma: the present value of expected income of each would be greatest if they cut only small holes in their own traps while others cut large holes in their. Small trap holes are a form of defection, and just as in the Public Goods Game it is the dominant strategy for a self-interested individual. But a shrimper might resist the temptation to defect if he were both public spirited towards the other fishers and sufficiently patient to value the future opportunities that they all would lose were he to use traps with smaller holes. Fehr and Andreas Leibbrandt implemented both a Public Goods Game and an experimental measure of impatience with the shrimpers. They found that both patience and cooperativeness in the game predicted larger trap holes (Fehr and Leibbrandt (2011)). The effects, controlling for a large number of other possible influences on hole size, were substantial. A shrimper whose experimentally measured patience and cooperativeness is a standard deviation greater than the mean is predicted to cut holes in his traps that are half a standard deviation larger than the mean.

Additional evidence of external validity comes from a set of experiments and field studies with 49 groups of herders of the Bale Oromo people in Ethiopia who were engaged in forest commons management. Devesh Rustagi and his coauthors implemented public goods experiments with a total of 679 herders. They also studied the success of the herders' cooperative forest projects. The most common behavioral type in the experiments, constituting a bit more than a third of the subjects, were "conditional cooperators" who responded positively to higher contributions by others. Controlling for a large number of other influences on the success of the forest projects, the authors found that groups with more conditional cooperators were more successful, in terms of number of new trees planted, than groups with fewer conditional cooperators. This was in part because members of groups with more conditional cooperators spent significantly more time monitoring the use of the forest by others. As in the case of the Brazilian shrimpers, the effects of group composition were large. A 10% increase in the fraction of experimentally identified conditional cooperators in a group was associated with an increase in trees planted or time spent monitoring by members of the group of about 3% (Rustagi, Engel and Kosfeld (2010)).

The available evidence suggests that students volunteering for experiments are not more pro-social in their orientations than other students (Falk, Meier and Zehnder (2011)); nor are student subjects more pro-social than non-students, indeed the reverse seems to be the case. (Fehr and List (2004), List (2004), Cardenas (2005), Carpenter, Verhoogen and Burks (2005), Bellemare, Kröger and Van Soest (2008), Carpenter, Connolly and Myers (2008), Burks, Carpenter and Goette (2009), Baran, et al. (2010), Cleave, Nikiforakis and Slonim (2010), Cardenas (2011), Falk, et al. (2011) and see Supplementary online material for a description of these studies.)

Levitt and List (2007) are right that care should be taken in generalizing experimental behavior to the real world. But none of the external validity concerns we have considered is sufficient to dismiss the experimental evidence that social preferences are important behavioral motivations and that these preferences may be affected by explicit incentives. This is especially the case when experimental subjects exhibit motives such as reciprocity, generosity and trust that allow a consistent explanation of otherwise anomalous real world examples of crowding in or out, such as those mentioned at the outset.

11. Optimal incentives for the sophisticated social planner

There are multiple plausible interpretations of the mechanisms underlying non-separability in the experiments we have presented, as is clear from the substantial size of the intersections among the hypothesized crowding out mechanisms that is evident in Figure 2. It would nonetheless be difficult, in light of these data, to sustain the implicit separability assumption adopted in many economic models.

A sophisticated social planner (or mechanism designer) – one who knows that the separability assumption is likely to be violated – faces a challenge that has yet to be addressed in the public economics literature: how to design optimal taxes, fines, or subsidies when the preferences that will determine citizen's responses depend on the incentives deployed. Thus, the designer must consider the effects – whether state-dependent or endogenous – of the instruments under consideration on individuals' social preferences and evaluate alternative policies on the basis of the resulting joint equilibrium of these preferences and economic allocations.

The problem facing the planner is quite a bit more difficult than the one we faced writing this paper. We studied the effects of incentives in experiments and natural settings and then sought ex post to determine the kinds of nonseparability – categorical or marginal crowding out or in – that might explain the results. The planner, however, must determine, ex ante whether the separability assumption is likely to be violated, and if so, how. The challenge is even greater because the nature and extent of non-separability itself is not given but (as we will see) may be influenced by the overall policy package of which the incentives are a part.

We begin with the more modest way of addressing the planner's problem and consider as exogenously given the nature and degree of the indirect effects of incentives on social preferences (that is, the signs and the size of the crowding parameters $\lambda_c, \lambda_m, \Lambda_c$ and Λ_m) Given the nature and extent of nonseparability, we then seek to determine the optimal level or mix of incentives taking account of their effects on preferences (Fershtman and Heifetz (2006), Heifetz, Segev and Talley (2007), Bowles and Hwang (2008), Hwang and Bowles (2011a)).

Here, two results may guide the social planner. The first is that in the presence of crowding out, incentives and social preferences are substitutes, so the deleterious indirect effect of incentives will be least where individual social preferences are modest or nonexistent (as will be the case in the endogenous preference model if there are few or no public spirited citizens or in the state-dependent model where the citizen's baseline social preferences are modest or zero). Societies in which social preferences are more prevalent not only may be able to afford less use of incentives but will find them less effective (when both direct and indirect effects are accounted for) than would be the case in a less civic minded culture. By the symmetry of the definition of substitutes (see equations 6 and 13) in the presence of crowding out, policies to enhance social preferences (that is raising γ or λ_0) will be more effective in promoting contributions the public good where incentives are little used.

In a cultural-institutional dynamic setting where economic incentives and socialization practices to promote civic mindedness are adopted as alternative measures to enhance public goods provision, this substitutability property of incentives and social preferences may support at least two evolutionarily stable equilibria. In one, extensive use of incentives is coupled with relatively low levels of civic mindedness in the population. In this state there is

little incentive to inculcate social preferences, the effect of which would be modest given the crowding phenomenon. In the other cultural-institutional equilibrium a social planner serving a civic minded population makes more modest use of incentives due to their limited effectiveness, once their crowding out effects are accounted for.

The second result for the social planner takes us back to Titmuss and others who concluded that if incentives crowd out social preferences then incentives will be overused by a naïve planner who is unaware of the effects of incentives on preferences. As a result, in these cases the sophisticated planner would either not use incentives, or would use them less than would the naïve planner. But the prescription that incentives are overused does not follow from the (correct) observation that crowding out occurs: it is readily shown that when crowding out occurs the sophisticated planner may make either greater or lesser use of explicit incentives than would her naïve counterpart (Bowles and Hwang (2008), Hwang and Bowles (2011a)).

The sophisticated planner may make greater use of incentives when incentives crowd out social preferences is that if incentives work less well than would be the case under separability, then there are two offsetting influences on their optimal use. The one that forms the basis of the Titmuss critique is that crowding out reduces the marginal effect of the subsidy on the target's behavior; and if this were the only effect Titmuss would be right. But there is a second often overlooked effect. Because the incentive is less effective (either categorically or marginally), the under provision of the public good will be exacerbated (compared to what would occur were crowding out absent) and if the benefits of the public good are concave in the amount provided the marginal benefit of altering the target's behavior is therefore correspondingly greater.

The intuition is transparent: the doctor who discovers that a treatment he has been prescribing is less effective than he thought may opt for stronger doses rather than weaker or for abandoning the treatment. As long as there are diminishing marginal returns to the public good and crowding out is categorical (and not too large) the naïve social planner will make too little use of the incentive. The reason is that in this case crowding does not change the marginal effect of the incentive on the citizens' contribution level; but the reduction in the public good resulting from crowding means that the marginal benefits to increasing its supply rise. (If categorical crowding is sufficiently large the naïve planner will over-use the incentive because the sophisticated planner will choose no incentive at all in this case.) But the sophisticated planner may make greater use of incentives even when only marginal crowding out occurs, if the benefit function is sufficiently concave.
A less modest approach to the design of appropriate incentives where separability may not hold is to recognize that the extent of the non-separability problem (that is, the magnitudes of the crowding parameters in the models of section 3 and 4, namely λ_c, λ_m, Λ_c, and Λ_m) is not exogenous, but can be affected by the nature of the incentives and the manner in which they are deployed. Designing policies that can convert incentives from being substitutes for social preferences to being their complements, however, requires an understanding of why crowding out occurs.

The most plausible explanation for the failure of the separability assumption is that when people engage in trade, produce goods and services, save and invest, they are not only attempting to get things, they are also trying to be someone, both in their own eyes and in the eyes of others (Cooley (1902), Leung and Martin (2003), Akerlof and Kranton (2010), Bloom (2010)). We refer to the second – the being or becoming motives – as constitutive. Incentives addressed to our acquisitive desires sometimes appear to dampen or impede the pursuit of our

constitutive aspirations. Among the reasons, we have seen, are that in addition to affecting the costs and benefits of an action, incentives also provide information about the person imposing the incentive, suggest appropriate behavior by framing decision situations, may compromise the target's sense of autonomy, and alter the environments in which we learn new preferences.

This may explain why incentives for settlement of conflicts may fail. Representative samples of Jewish West Bank settlers in 2005, Palestinian refugees in 2005, and Palestinian students in 2006 were asked how angry and disgusted they would feel or how supportive of violence they might be if their political leaders were to compromise on contested issues between the groups. Those who regarded their group's claims (on Jerusalem, for example) as reflecting "sacred values" (about half in each of the three groups) expressed far greater anger, disgust and support for violence if the compromise were accompanied by a monetary compensation for their own group than if no compensation were offered (Ginges, Atran, Medin, et al. (2007)). Similar results were fund in a survey of the willingness of Swiss citizens to accept environmental hazards (Frey and Oberholzer-Gee (1997)). (For a discussion on environmental motivation and crowding effects see Frey and Stutzer (2008).)

The importance of constitutive rather than acquisitive motives may be at work in the negative response to incentives that convey adverse information about the individual imposing the incentives. Recall that in the Trust Game implemented by Fehr and Rockenbach (2003) the investor's threat to fine the trustee if the back transfer was not sufficient resulted in a lesser level of reciprocity of the trustee: conditional on the investor's transfer to the trustee, back-transfers were less under the fine condition. This was especially the case when it appeared that the intent of the fine was to induce the trustee to grant most of the joint surplus to the investor. Where the announced desired back-transfer would have allowed the investor to capture most of the surplus had the trustee complied, use of the threat of the fine was very counter productive.

The back-transfers from those facing an investor who declined use the fine averaged a generous 60 percent of what the trustee could have kept; while those threatened with a fine transferred just 22 percent (almost half gave nothing and paid the fine!) But where the investor announced a desired back transfer such that the trustee who complied would gain half or more of the joint surplus, back transfers from those threatened with fine for non compliance were lower (35 percent) but not significantly so than among those facing an investor who chose not to use the fine (43 percent.) It appears that the use of the fine in these conditions signaled the unfair intent of the investor, rather than simply his distrust of the trustee.

Interpretations other than this "unfair intent" result are possible, however, for the larger was the desired back transfer, the more costly was compliance. Thus for larger demands, simply returning nothing and paying the fine (as many of the subjects did) might have been attractive to self-interested subjects, who had they been faced with a lower demand would have maximized payoffs by complying. It seems from this and similar experiments that fines may have negative effects even if imposed to implement a fair outcome and even when the decision to use the fine was not made by the investor, but rather by chance (Fehr and List (2004), Houser, Xiao, McCabe, et al. (2008b)). But the experiments also are consistent with the idea that threats deployed in self-interested ways can backfire.

This helps explain the very different effect of incentives imposed by peers who do not stand

to benefit personally. An example is the Public Goods experiment in which fellow group members have the opportunity to reduce their own payoffs in order to punish (reduce the payoffs of) others in their group once each member's contributions are revealed (Fehr and Gächter (2000), (2002a), Masclet, et al. (2003)). One treatment in these public goods experiments is particularly revealing: group membership is shuffled after each period so that in subsequent periods a punisher will not be in the same group with the target of his or her punishment, and thus cannot benefit from the target's response. Punishment in this case is an altruistic act as it benefits others at the expense of the punisher and hence it cannot be interpreted as a signal of unfair intent. In this setting there is a strong positive response by low contributors.

Although there is no direct evidence, a plausible explanation of the effectiveness of incentives in this case is that when punished by a peer who had nothing to gain by doing so, those who have contributed less than others interpret the punishment as a signal of public-spirited social disapproval by fellow group members seeking to uphold a social norm and willing to sacrifice payoffs to do so. As a result, targeted free riders and even free riders who escaped punishment feel shame, which they redress by subsequently contributing more. In this case the incentive (prospect of peer imposed fines) has crowded in social preferences. These experiments illustrate the opposite of the "bad news about the principal" results in section 5. The principals here are the peers who punish free riding fellow group members; and the positive response to the fines in this case may reflect the fact that the willingness to pay to punish defectors with no expectation of personal gain is good news about the person implementing the incentive.

Consistent with the interpretation that crowding out does not follow from the use of incentives per se, but rather from the meaning that the incentives convey to the participants is an extension of the "control aversion" experiment of Falk and Kosfeld (2006) described in section 5. Schnedler and Vadovic (2011) found that when agents themselves implemented controls (rather than the principal) the negative response did not occur. A large number of experiments have found positive effects of incentives imposed by the decision of the targets of the incentives rather than by the experimenter or by a principal (Cardenas (2005), Tyran and Feld (2006), Kroll, Cherry and Shogren (2007), Ertan, Page and Putterman (2009), Kosfeld, Okada and Riedl (2009), Mellizo, Carpenter and Matthews (2011), Sutter, Haigner and Kocher (2011)).

John Stuart Mill (whose definition of the restrictive boundaries of our discipline we mentioned at the outset) and economists since have recognized that the purposes of individual economic action are constitutive as well as acquisitive (Akerlof and Kranton (2010)). But what some have missed is that our acquisitive and constitutive motivations may not be separable.

Some of the founders of economics knew this. Jeremy Bentham's *Introduction to the Principles of Morals and Legislation* (1789), is arguably the first text in what we now call public economics. In it he explained how proper incentives should harness self-interested objectives for public ends by making "it each man's interest to observe ... that conduct which it is his duty to observe." In other words, make sure that doing his duty is incentive compatible.

But he also understood the constitutive side of action and the need to design incentives that are complements of the moral sentiments rather than substitutes:

> A punishment may be said to be …a moral lesson, when by reason of the ignominy it stamps upon the offence, it is calculated to inspire the public with sentiments of aversion towards those pernicious habits and dispositions with which the offence appears to be connected; and thereby to inculcate the opposite beneficial habits and dispositions (Bentham (1970 [1789]): p.26).

Few economists followed Bentham in this. An exception is Albert Hirschman, who noted that economists seek to deal with unethical or antisocial behavior by raising the cost of that behavior rather than proclaiming standards and imposing prohibitions and sanctions. The reason is probably that they think of citizens as consumers with unchanging or arbitrarily changing tastes in matters civic as well as commodity-related behavior. . . A principal purpose of publicly proclaimed laws and regulations is to stigmatize antisocial behavior and thereby to influence citizens' values and behavioral codes (Hirschman (1985): p. 10).
The fact that punishments are not only incentives but also "moral lessons" that "stigmatize antisocial behavior" may help resolve one of the puzzles in the literature we have just surveyed. In a widely cited natural experiment, the imposition of fines on parents arriving late to pick up their children at day care centers in Haifa resulted in a doubling of the number of tardy pickups (Gneezy and Rustichini 2000a). But the small tax on plastic grocery bags enacted in Ireland in 2002 had the opposite effect: in two weeks it resulted in a 94 percent decline in their use and appeared to crowd in social preferences (Rosenthal (2008)).

The contrast is instructive. In the Haifa case, the experimenters (respecting standard experimental protocols) provided no justification for the introduction of the fine on the tardy parents. Moreover the parents' occasional lateness could have occurred for reasons beyond their control, rather than as the result of a deliberate disregard for the inconvenience it caused the teachers. Finally, lateness was not so common as to be widely broadcast to the other parents. By contrast, the introduction of the Irish plastic bag tax was preceded by a substantial publicity campaign, and the use of the bags required a deliberate choice made in a highly public condition. In the Irish case, as in the experiment by Vertova and Galbiati (2010) mentioned in section 9, the monetary incentive was introduced jointly with a message of explicit social obligation, and it apparently served as a reminder of the larger social costs of the use and disposal of the bags.

The same message comes from a voting study. In Switzerland, the removal of a negligible fine for not voting significantly reduced voting turnout; but a considerable reduction in the cost of voting (by allowing balloting by mail) had no effect on turnout. The implication is that the fine for not voting encouraged turnout not as an incentive (by affecting the costs of not voting) but rather as a message of the importance of one's civic duty (Funk (2007)).

The fact that fines often work more as messages than as incentives poses a problem for the sophisticated planner because the same intervention may bear radically different messages in different cultures. Bohnet and her coauthors implemented a Trust Game in which in one treatment the investor had the option of reducing the payoffs of trustees who betrayed their trust (Bohnet et al. (2010).) Compared to the treatment in which this so-called "revenge" option was not available, when they had the revenge option a substantially larger fraction of Saudi investors trusted their partner, while a substantially smaller fraction of American investors trusted. Making trust more incentive compatible thus had diametrically opposed effects in the two cultures.

12. Conclusion: Are Incentives to Blame?

Is there a simple lesson for public policy? We think there is. Titmuss was right that incentives sometimes crowd out noneconomic motives, and this may degrade economic performance. But Titmuss and the literature that followed him targeted incentives per se as the cause of crowding out and recommended a reduced role for incentives in the governance of economic interactions.

Both the diagnosis and the policy implication are wrong. Crowding out, as we have seen, may require greater, not lesser use of incentives. And perhaps more important: fines, subsidies, and other monetary incentives per se may not be the culprit. What accounts for crowding out, we believe, is the meaning of the fines or subsidies to the target of the incentives; and this depends on the social relationships among the actors, the information the incentive provides, and the preexisting normative frameworks of the actors. This is the message of the contrast between the Irish grocery bag tax and the Haifa fines for tardiness, along with the fact that fines imposed on low contributors by peers in Public Goods Games have positive effects while fines imposed by principals on agents sometimes backfire. In addition, incentives chosen by agents themselves (for example, by majority rule of team members), may have a more positive effect on individual performance than if they are imposed (Mellizo, Carpenter, and Matthews (2011)).

Fines deployed either to exploit or to control the target (or that give this appearance or that have this effect) are likely to be less effective than they would under separability and may even be counterproductive. The reason, we think, is that they activate the target's desire to constitute himself or herself as a dignified and autonomous individual who is treated fairly by others. It is this constitutive motive that sometimes trumps the acquisitiveness tapped by the incentive, and that leads to a contrary response. The same incentives deployed by individuals who do not stand to benefit personally, and that are intended to foster pro-social behavior are more likely to be complements rather than substitutes for social preferences, crowding them in rather than out. They do this by activating rather than diminishing the target's constitutive motives such as the desire to be treated fairly and to treat others fairly, to be a good member of a community, and the feeling of shame when others regard one as having failed in this.

Present experimental and other evidence give insufficient guidance to the planner who wishes to know ex ante, the effects of the incentives he is considering implementing. But on the basis of what we do know a good rule might be the following: The policy package of which the incentives are part should let the target understand that the desired modification in her actions will serve to implement an outcome that is socially beneficial so that that the target is more likely to endorse the purpose of the incentive, rather than being offended by it as either unjust or a threat to her autonomy or in some other way reflecting badly on the intentions of the planner.

13. References

Aaron, Henry J. . 1994. "Distinguished Lecture on Economics in Government: Public Policy, Values, and Consciousness." Journal of Economic Perspectives, 8(2), pp. 3-21.

Akerlof, George A. 1984. An Economic Theorist's Book of Tales. Cambridge, UK: Cambridge University Press.

Akerlof, George A. and Rachel Kranton. 2010. Identity Economics: How Our Identities

Shape Our Work, Wages, and Well-Being. Princeton: Princeton University Press.
Al-Ubaydli, Omar; Daniel Houser; John Nye; Maria Pia Paganelli and Xiaofei (Sophia) Pan. 2011. "The Causal Effect of Market Participation on Trust: An Experimental Investigation Using Randomized Control " In Mimeo.
Andreoni, James. 1990. "Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving." The Economic Journal, 100(401), pp. 464-77.
Angrist, Joshua and Victor Lavy. 2009. "The Effects of High Stakes High School Achievement Rewards: Evidence from a Randomized Trial." American Economic Review, 99(4), pp. 1384-414.
Ariely, D.; U. Gneezy; G. Loewenstein and N. Mazar. 2009. "Large Stakes and Big Mistakes." Review of Economic Studies, 76(2), pp. 451-69.
Ariely, Dan; Anat Bracha and Stephan Meier. 2009. "Doing Good or Doing Well? Image Motivation and Monetary Incentives in Behaving Prosocially." American Economic Review, 99(1), pp. 544-55.
Arrow, Kenneth J. 1972. "Gifts and Exchanges." Philosophy and Public Affairs, 1(4), pp. 343-62.
____. 1971. "Political and Economic Evaluation of Social Effects and Externalities," In Frontiers of Quantitative Economics, ed. M. D. Intriligator, 3-23. Amsterdam: North Holland.
Bandiera, Oriana; Iwan Barankay and Imran Rasul. 2005. "Social Preferences and the Response to Incentives: Evidence from Personnel Data." The Quarterly Journal of Economics, 120(3), pp. 917-62.
Bandura, Albert. 1991. "Social Congnitive Theory of Moral Thought and Action," In Handbook of Moral Behavior and Development: Volume I, Theory, ed. W. Kurtines and J. Gewirtz, 45-103. Hillsdale, New Jersey: Lawrence Erlbaum and Associates.
Bar-Gill, Oren and Chaim Fershtman. 2005. "Public Policy with Endogenous Preferences." Journal of Public Economic Theory, 7(5), pp. 841-57.
Baran, Nicole M.; Paola Sapienza and Luigi Zingales. 2010. "Can We Infer Social Preferences from the Lab? Evidence from the Trust Game," In NBER Working Papers.
Barr, Abigail. 2004. "The Effects of Birthplace and Current Context on Other Regarding Preferences in Accra," Forthcoming n Experimenting with Social Norms: Fairness and Punishment in Cross-Cultural Perspective, ed. J. Ensminger and J. Henrich. unpublished manuscript submitted: The Russell Sage Foundation Press.
____. 2001. "Social Dilemmas, Shame-Based Sanctions, and Shamelessness: Experimental Results from Rural Zimbabwe," In. Oxford University: Centre for the Study of African Economies Working Paper WPS/2001.11.
Barr, Abigail; Chris Wallace; Jean Ensminger; Joseph Henrich; Clark Barrett; Alexander Bolyanatz; Juan Camilo Cardenas; Michael Gurven; Edwins Gwako; Carolyn Lesorogol, et al. 2009. "Homo Æqualis: A Cross-Society Experimental Analysis of Three Bargaining Games," In Discussion Paper Series. Oxford: Department of Economics, University of Oxford.
Becker, Gary S. 1976. "Altruism, Egoism, and Genetic Fitness: Economics and Sociobiology." Journal of Economic Literature, 14(3), pp. 817-26.
Belkin, Douglas. 2002. "Boston Firefighters Sick - or Tired of Working 15-Day Allowance Seen Fueling Call-Ins," In Boston Globe, B1. Boston.
Bellemare, Charles; Sabine Kröger and Arthur Van Soest. 2008. "Measuring Inequity Aversion in a Heterogeneous Population Using Experimental Decisions and Subjective Probabilities." Econometrica, 76(4), pp. 815-39.
Benabou, Roland and Jean Tirole. 2006. "Incentives and Prosocial Behavior." American Economic Review, 96(5), pp. 1652-78.
____. 2003. "Intrinsic and Extrinsic Motivation." Review of Economic Studies, 70, pp.

489-520.
Bentham, Jeremy. 1970 [1789]. An Introduction to the Principles of Morals and Legislation. Athlone Press.
Benz, Matthias and Stephan Meier. 2008. "Do People Behave in Experiments As in the field?—Evidence from Donations." Experimental Economics, 11(3), pp. 268-81.
Besley, T.; G. Bevan and K. Burchardi. 2009. "Naming & Shaming: The Impacts of Different Regimes on Hospital Waiting Times in England and Wales'," In CEPR Discussion Paper no. 7306. London: Centre for Economic Policy Research. .
Bewley, Truman F. 1999. Why Wages Don't Fall During a Recession. Cambridge: Harvard University Press.
Birch, L L and D W Marlin. 1982. "I Don't Like It; I Never Tried It: Effects of Exposure on Two-Year-Old Children's Food Preferences." Appetite, 3(4), pp. 353-60.
Bisin, Alberto and Thierry Verdier. 2001. "The Economics of Cultural Transmission and the Dynamics of Preferences." Journal of Economc Theory, 97(2), pp. 298-319.
Bisin, Alberto and Thierry Verdier. 2011. "The Economics of Cultural Transmission and Socialization," In Handbook of Social Economics, ed. J. Benhabib, A. Bisin and M. O. Jackson, 339-416. The Netherlands: North-Holland Elsevier.
Bliss, Christopher J. 1972. "Review of R.M. Titmuss, the Gift Relationship: From Human Blood to Social Policy." Journal of Public Economics, 1, pp. 162-65.
Bloom, Paul. 2010. How Pleasure Works: The New Science of Why We Like What We Like. New York: Norton.
Bohnet, I and Y Baytelman. 2007. "Institution and Trust- Implications for Preferences, Beliefs, and Behavior." Rationality and Society, 19(1), pp. 99-135.
Bohnet, Iris; Bruno Frey and Steffen Huck. 2001. "More Order with Less Law: On Contractual Enforcement, Trust, and Crowding." American Political Science Review, 95(1), pp. 131-44.
Bohnet, Iris; Fiona Greig; Benedikt Herrmann and Richard Zeckhauser. 2008. "Betrayal Aversion: Evidence from Brazil, China, Oman, Switzerland, Turkey, and the United States." American Economic Review, 98(1), pp. 294-310.
Bohnet, Iris; Benedikt Herrmann; Mohamad Al-Ississ; Andrea Robbett; Khalid Al-Yahia and Richard Zeckhauser. 2010. "The Elasticity of Trust: How to Promote Trust in the Arab Middle East and the United States," In Faculty Research Working Paper Series. Cambridge: Harvard Kennedy School.
Borges, Goerg and Bernd Irlenbusch. 2007. "Fairness Crowded out by Law: An Experimental Study of Withdrawal Rights." Journal of Instittional and Theoretical Economics, 163, pp. 84-101.
Bowles, S. and H. Gintis. 2011. A Cooperative Species: Human Reciprocity and Its Evolution. Princeton: Princeton University Press.
Bowles, Samuel. 1998. "Endogenous Preferences: The Cultural Consequences of Markets and Other Economic Institutions." Journal of Economic Literature, 36(1), pp. 75-111.
____. 2011. "Is Liberal Society a Parasite on Tradition?" Philosophy & Public Affairs, 39(1), pp. 46-81.
____. 1989. "Mandeville's Mistake: Markets and the Evolution of Cooperation," In Presented to the September Seminar, University College London. London.
____. 2004. Microeconomics: Behavior, Institutions, and Evolution. Princeton: Princeton University Press.
____. 2008. "Policies Designed for Self-Interested Citizens May Undermine "the Moral Sentiments": Evidence from Economic Experiments." Science, 320(5883), pp. 1605-09.
Bowles, Samuel and Herbert Gintis. 2005. "Prosocial Emotions," In The Economy as a

Complex Evolving System Iii: Current Perspectives and Future Directions, ed. S. Durlauf and L. Blume, 339-66. New York: Oxford University Press.

Bowles, Samuel and Sung-Ha Hwang. 2008. "Social Preferences and Public Economics: Mechanism Design When Social Preferences Depend on Incentives." Journal of Public Economics, 92(8-9), pp. 1811-20.

Boyd, Robert and Peter J. Richerson. 1985. Culture and the Evolutionary Process. Chicago: University of Chicago Press.

Burks, Stephen; Jeffrey Carpenter and Lorenz Goette. 2009. "Performance Pay and Worker Cooperation: Evidence from an Artefactual Field Experiment." Journal of Economic Behavior & Organization, 70(3), pp. 458-69.

Cardenas, Juan. 2011. "Social Norms and Behavior in the Local Commons as Seen through the Lens of Field Experiments." Environmental and Resource Economics, pp. 1-35.

Cardenas, Juan Camilo. 2005. "Groups, Commons and Regulations: Experiments with Villagers and Students in Colombia," In Psychology, Rationality and Economic Behaviour: Challenging Standard Assumptions, ed. B. Agarwal and A. Vercelli, 242-70. International Economics Association.

____. 2004. "Norms from Outside and Inside: An Experimental Analysis on the Governance of Local Ecosystems." Forest Policy and Economics, 6, pp. 229-41.

Cardenas, Juan Camilo; John K. Stranlund and Cleve E. Willis. 2000. "Local Environmental Control and Institutional Crowding-Out." World Development, 28(10), pp. 1719-33.

Carpenter, Jeffrey; Samuel Bowles; Herbert Gintis and Sung-Ha Hwang. 2009. "Strong Reciprocity and Team Production: Theory and Evidence." Journal of Economic Behavior and Organization, In press.

Carpenter, Jeffrey; Cristina Connolly and Caitlin Myers. 2008. "Altruistic Behavior in a Representative Dictator Experiment." Experimental Economics, 11(3), pp. 282-98.

Carpenter, Jeffrey and Caitlin Knowles Myers. 2010. "Why Volunteer? Evidence on the Role of Altruism, Image, and Incentives." Journal of Public Economics, 94(11-12), pp. 911-20.

Carpenter, Jeffrey P. and Erika Seki. 2010. "Do Social Preferences Increase Productivity? Field Experimental Evidence from Fishermen in Toyama Bay." Economic Inquiry, in press.

Carpenter, Jeffrey; Eric Verhoogen and Stephen Burks. 2005. "Comparing Students to Workers: The Effects of Social Framing on Behavior in Distribution Games," In Research in Experimental Economics Series, 393-98.

Cavalli-Sforza, L. L. and Marcus W. Feldman. 1981. Cultural Transmission and Evolution : A Quantitative Approach. Princeton, N.J.: Princeton University Press.

Cervellati, Matteo; Joan Esteban and Laurence Kranich. 2010. "Work Values, Endogenous Sentiments Redistribution." Journal of Public Economics, 94(9-10), pp. 612-27.

Cleave, Blair L.; Nikos Nikiforakis and Robert Slonim. 2010. "Is There Selection Bias in Laboratory Experiments?," In Economics Working Paper Series. Sidney: University of Sidney.

Cohn, Alain; Ernst Fehr and Lorenz Goette. 2011. "Fair Wages and Effort: Evidence from a Field Experiment," In Mimeo.

Cooley, C. H. 1902. Human Nature and Social Order. New York: Scribner's.

Cooter, Robert. 2000. "Do Good Laws Make Good Citizens? An Economic Analysis of Internalized Norms." Virginia Law Review, 86(8), pp. 1577-601.

Deci, Edward L. 1975. Intrinsic Motivation. New York: Plenum.

Deci, Edward L.; Richard Koestner and Richard M. Ryan. 1999. "A Meta-Analytic Review of Experiments Examining the Effects of Extrinsic Rewards on Intrinsic Motivation."

Psychological Bulletin, 125(6), pp. 627-68.
Deci, Edward L. and Richard M. Ryan. 1985. Intrinsic Motivation and Self-Determination in Human Behavior. New York and London: Plenum Press.
DellaVigna, Stefano. 2009. "Psychology and Economics: Evidence from the Field." Journal of Economic Literature, 47(2), pp. 315-72.
Dickenson, David and Marie-Claire Villeval. 2008. "Does Monitoring Decrease Work Effort? The Complementarity between Agency and Crowding-out Theories." Games and Economic Behavior, 63(1), pp. 56-76.
Drago, Francesco; Roberto Galbiati and Pietro Vertova. 2009. "The Deterrent Effects of Prison: Evidence from a Natural Experiment." Journal of Political Economy, 117(2), pp. 257-80.
Dufwenberg, Martin; Paul Heidhues; Georg Kirchsteiger; Frank Riedel and Joel Sobel. 2011. "Other-Regarding Preferences in General Equilibrium " Review of Economic Studies, (Published online: February 4, 2011).
Ellingsen, Tore and Magnus Johannesson. 2008. "Pride and Prejudice: The Human Side of Incentive Theory." American Economic Review, 98, pp. 990-1008.
Ellingsen, Tore; Magnus Johannesson; Johanna Möllerström and Sara Munkhammar. 2011. "Social Framing Effects: Preferences or Beliefs? ," In. Stockholm School of Economics, Ramböll Management and Harvard University. At: http://www2.hhs.se/personal/ellingsen/pdf/FramingGEB1.pdf.
Elster, Jon. 1998. "Emotions and Economic Theory." Journal of Economic Literature, 36(1), pp. 47-74.
Ertan, Arhan; Talbot Page and Louis Putterman. 2009. "Who to Punish? Individual Decisions and Majority Rule in Mitigating the Free Rider Problem." European Economic Review, 53(5), pp. 495-511.
Falk, Armin; Ernst Fehr and Urs Fischbacher. 2005. "Driving Forces Behind Informal Sanctions." Econometrica, 73(6), pp. 2017-30.
Falk, Armin; Ernst Fehr and Christian Zehnder. 2006. "Fairness Perceptions and Reservation Wages -- the Behavioral Effects of Minimum Wage Laws." Quarterly Journal of Economics, (1347-1381).
Falk, Armin; Simon Gächter and Judit Kovacs. 1999. "Intrinsic Motivation and Extrinsic Incentives in a Repeated Game with Incomplete Contracts." Journal of Economic Psychology, 20, pp. 251-64.
Falk, Armin and James J. Heckman. 2009. "Lab Experiments Are a Major Source of Knowledge in the Social Sciences." Science, 326(5952), pp. 535-38.
Falk, Armin and Michael Kosfeld. 2006. "The Hidden Costs of Control." American Economic Review, 96(5), pp. 1611-30.
Falk, Armin; Stephan Meier and Christian Zehnder. 2011. "Did We Overestimate the Role of Social Preferences? The Case of Self-Selected Student Samples," In IZA Discussion Papers, ed. I. f. t. S. o. L. (IZA).
Falkinger, Josef. 1996. "Efficient Private Provision of Public Goods by Rewarding Deviations from Average." Journal of Public Economics, 62(3), pp. 413-22.
Falkinger, Josef; Ernst Fehr; Simon Gächter and Rudolf Winter-Ebmer. 2000. "A Simple Mechanism for the Efficient Provision of Public Goods." American Economic Review, 90(1), pp. 247-64.
Fehr, Ernst and Simon Gächter. 2002a. "Altruistic Punishment in Humans." Nature, 415, pp. 137-40.
____. 2000. "Cooperation and Punishment in Public Goods Games." American Economic Review, 90(4), pp. 980-94.
____. 2002b. "Do Incentive Contracts Crowd out Voluntary Cooperation?," In. Institute

for Empirical Research in Economics. University of Zurich. Working Paper Series.
Fehr, Ernst; Simon Gächter and Georg Kirchsteiger. 1997. "Reciprocity as a Contract Enforcement Device: Experimental Evidence." Econometrica, 65(4), pp. 833-60.
Fehr, Ernst and Lorenz Goette. 2007. "Do Workers Work More If Wages Are High? Evidence from a Randomized Field Experiment." American Economic Review, 97(1), pp. 298.
Fehr, Ernst; Alexander Klein and Klaus M. Schmidt. 2007. "Fairness and Contract Design." Econometrica, 75(1), pp. 121-54.
Fehr, Ernst and Andreas Leibbrandt. 2011. "A Field Study on Cooperativeness and Impatience in the Tragedy of the Commons." Journal of Public Economics, 95(9–10), pp. 1144-55.
Fehr, Ernst and John List. 2004. "The Hidden Costs and Returns of Incentives: Trust and Trustworthiness among Ceos." Journal of The European Economic Association, 2(5), pp. 743-71.
Fehr, Ernst and Bettina Rockenbach. 2003. "Detrimental Effects of Sanctions on Human Altruism." Nature, 422(13 March), pp. 137-40.
Fehr, Ernst and Klaus M. Schmidt. 2007. "Adding a Stick to the Carrot? The Interaction of Bonuses and Fines." American Economic Review, 97(2), pp. 177-81.
____. 1999. "A Theory of Fairness, Competition, and Cooperation." Quarterly Journal of Economics, 114(3), pp. 817-68.
Fershtman, Chaim and Aviad Heifetz. 2006. "Read My Lips, Watch for Leaps: Preference Equilibrium and Political Instability." The Economic Journal, 116, pp. 246-65.
Fiske, Alan Page. 1992. "The Four Elementary Forms of Sociality: Framework for a Unified Theory of Social Relations." Psychological Review, 99(4), pp. 689-723.
Frey, B. and R Jegen. 2001. "Motivation Crowding Theory: A Survey of Empirical Evidence." Journal of Economic Surveys, 15(5), pp. 589 - 611.
Frey, Bruno S. 1997. "A Constitution for Knaves Crowds out Civic Virtues." Economic Journal, 107(443), pp. 1043-53.
____. 1993. "Does Monitoring Increase Work Effort? The Rivalry with Trust and Loyalty." Economic Inquiry, 31, pp. 663-70.
Frey, Bruno S. and Felix Oberholzer-Gee. 1997. "The Cost of Price Incentives: An Empirical Analysis of Motivation Crowding-Out." American Economic Review, 87(4), pp. 746-55.
Frey, Bruno S. and Alois Stutzer. 2008. "Environmental Morale and Motivation," In The Cambridge Handbook of Psychology and Economic Behaviour, ed. A. Lewis. Cambridge: Cambridge University Press.
Fryer, Roland. 2011. "Financial Incentives and Student Acheivement: Evidence from Randomized Trials." Quarterly Journal of Economics, 126(4).
Funk, Patricia. 2007. "Is There an Expressive Function of Law? An Empirical Analysis of Voting Laws with Symbolic Fines." American Law and Economics Review.
Gächter, Simon and Armin Falk. 2002. "Reputation or Reciprocity? Consequences for Labour Relation." Scandinavian Journal of Economics, 104(1), pp. 1 - 26.
Gächter, Simon and Ernst Fehr. 1999. "Collective Action as a Social Exchange." Journal of Economic Behavior and Organization, 39(4), pp. 341-69.
Gächter, Simon; Esther Kessler and Manfred Konigstein. 2011. "The roles of incentives and voluntary cooperation for contractual compliance" In University of Nottingham, School of Economics. Nottingham: CeDEx Discussion Paper Series no. 2011-06
Gächter, Simon; Daniele Nosenzo and Martin Sefton. 2010. "The Impact of Social Comparisons on Reciprocity," In Discussion Papers. The Centre for Decision Research and Experimental Economics, School of Economics, University of Nottingham. Forthcoming in

Scandinavian Journal of Economics.
Galbiati, Roberto and Pietro Vertova. 2008. "Obligations and Cooperative Behaviour in Public Good Games." Games and Economic Behavior, 64(1), pp. 146-70.
Ginges, J; Scott Atran; Douglas Medin and Khalil Shikaki. 2007. "Sacred Bounds on Rational Resolution of Violent Political Conflict." Proceedings of the National Acadamy of Science, 104(18), pp. 7357-60.
Gneezy, U.; S. Meier and P. Rey-Biel. 2011. "When and Why Incentives (Don't) Work to Modify Behavior." The Journal of Economic Perspectives, 25(4), pp. 191-209.
Gneezy, Uri. 2003. "The W Effect of Incentives," In. University of Chicago Graduate School of Business.
Gneezy, Uri and Aldo Rustichini. 2000a. "A Fine Is a Price." Journal of Legal Studies, 29(1), pp. 1-17.
____. 2000b. "Pay Enough or Don't Pay at All." Quarterly Journal of Economics, 115(2), pp. 791-810.
Grant, Ruth. 2011. Strings Attached: Untangling the Ethics of Incentives. . Princeton.
Greenberger, Scott. 2003. "Sick Day Abuses Focus of Fire Talks," In Boston Globe, B7.
Guth, Werner and Harmutt Kliemt. 1994. "Competition or Co-Operation: On the Evolutionary Economics of Trust, Exploitation, and Moral Attitudes." Metroeconomica, 45(2), pp. 155-87.
Healy, Kieran. 2006. Best Last Gifts: Altruism and the Market for Human Blood and Organs. Chicago: University of Chicago Press.
Heifetz, Aviad; Ella Segev and Eric Talley. 2007. "Market Design with Endogenous Preferences." Games and Economic Behavior, 58, pp. 121-53.
Helsley, Robert W. and William C. Strange. 2000. "Social Interactions and the Institutions of Local Government." The American Economic Review, 90(5), pp. 1477-90.
Henrich, Joe; Robert Boyd; Samuel Bowles; Colin Camerer; Ernst Fehr; Herbert Gintis; Richard McElreath; Michael Alvard; Abigail Barr; Jean Ensminger, et al. 2005. "'Economic Man' in Cross-Cultural Perspective: Behavioral Experiments in 15 Small-Scale Societies." Behavioral and Brain Sciences, 28, pp. 795-855.
Henrich, Joseph; Jean Ensminger; Richard McElreath; Abigail Barr; Clark Barrett; Alexander Bolyanatz; Juan Camilo Cardenas; Michael Gurven; Edwins Gwako; Natalie Henrich, et al. 2010. "Markets, Religion, Community Size, and the Evolution of Fairness and Punishment." Science, 327(5972), pp. 1480-84.
Herrmann, Benedikt and Henrik Orzen. 2008. "The Appearance of Homo Rivalis: Social Preferences and the Nature of Rent Seeking," In Discussion Papers 2008-10. Nottingham: The Centre for Decision Research and Experimental Economics, School of Economics, University of Nottingham.
Herrmann, Benedikt; Christian Thoni and Simon Gächter. 2008a. "Antisocial Punishment across Societies." Science, 319(7 March 2008), pp. 1362-67.
Heyman, James and Dan Ariely. 2004. "Effort for Payment: A Tale of Two Markets." Psychological Science, 15(11), pp. pp. 787-93.
Hirschman, Albert O. 1985. "Against Parsimony:Three Ways of Complicating Some Categories of Economic Discourse." Economics and Philosophy, 1(1), pp. 7-21.
____. 1977. The Passions and the Interests : Political Arguments for Capitalism before Its Triumph. Princeton, N.J.: Princeton University Press.
Hoffman, Elizabeth; Kevin McCabe; Keith Shachat and Vernon L. Smith. 1994. "Preferences, Property Rights, and Anonymity in Bargaining Games." Games and Economic Behavior, 7(3), pp. 346-80.
Holmås, Tor Helge; Egil Kjerstad; Hilde Lurås and Odd Rune Straume. 2010. "Does Monetary Punishment Crowd out Pro-Social Motivation? A Natural Experiment on Hospital

Length of Stay." Journal of Economic Behavior & Organization, 75(2), pp. 261-67.
Hopfensitz, Astrid and Ernesto Reuben. 2009. "The Importance of Emotions for the Effectiveness of Social Punishment*." The Economic Journal, 119(540), pp. 1534-59.
Houser, Daniel; Erte Xiao; Kevin McCabe and Vernon Smith. 2008a. "When Punishment Fails: Research on Sanctions, Intentions, and Non-Cooperation." Games and Economic Behavior, 62(2), pp. 509-32.
Hwang, Sung-Ha and Samuel Bowles. 2011a. "Optimal Incentives with State-Dependent Preferences," In Mimeo. Santa Fe Institute.
____. 2011b. "The Sophisticated Planner's Dillemma: Mechanism Design with Endogenous Preferences," In Santa Fe Institute. Santa Fe.
Irlenbusch, Bernd and Gabriele K Ruchala. 2008. "Relative Rewards within Team-Based Compensation." Labour Economics, 15 pp. 141-67.
Irlenbusch, Bernd and Dirk Sliwka. 2005. "Incentives, Decision Frames and Motivation Crowding out- an Experimental Investigation," In IZA Discussion paper No 1758 http://ssn.com/abstract=822866.
Karlan, Dean. 2005. "Using Experimental Economics to Measure Social Capital and Predict Financial Decisions." American Economic Review, 95(5), pp. 1688-99.
Kessler, Esther. 2008. "Behavioural Economics of Performance Incentives," In School of Economics,PhD Economics. Nottingham: University of Nottingham.
Kosfeld, Michael; Akira Okada and Arno Riedl. 2009. "Institution Formation in Public Goods Games." American Economic Review, 99(4), pp. 1335-55.
Kreps, David M. 1997. "Intrinsic Motivation and Extrinsic Incentives." The American Economic Review, 87(2), pp. 359-64.
Kroll, Stephan; Todd L. Cherry and Jason F. Shogren. 2007. "Voting, Punishment and Public Goods." Economic Inquiry, 45(3), pp. 557-70.
Laffont, Jean Jacques and Mohamed Salah Matoussi. 1995. "Moral Hazard, Financial Constraints, and Share Cropping in El Oulja." Review of Economic Studies, 62(3), p. 381-99.
Lazear, Edward. 2000. "Performance Pay and Productivity." American Economic Review, 90(5), pp. 1346 - 61.
Leibbrandt, Andreas; Uri Gneezy and John List. 2010. "Ode to the Sea: The Socio-Ecological Underpinnings of Social Norms," In University of Chicago, Department of Economics.
Leider, Stephen; Markus M. Möbius; Tanya Rosenblat and Quoc-Anh Do. 2009. "Directed Altruism and Enforced Reciprocity in Social Networks." The Quarterly Journal of Economics, 124(4), pp. 1815-51.
Lepper, Mark R.; Gerald Sagotsky; Janet L. Dafoe and David Greene. 1982. "Consequences of Superfluous Social Constraints: Effects on Young Children's Social Inferences and Subsequent Intrinsic Interest." Journal of Personality and Social Psychology, 42(1), pp. 51-65.
Leung, King To and John Levi Martin. 2003. "The Looking Glass Self: An Empirical Test and Elaboration." Social Forces, 81(3), pp. 843-79.
Levine, David K. 1998. "Modeling Altruism and Spitefulness in Experiments." Review of Economic Dynamics, 1(3), pp. 593-622.
Levitt, Steven D. and John List. 2007. "What Do Laboratory Experiments Measuring Social Preferences Reveal About the Real Word." Journal of Economic Perspectives, 21(1), pp. 153-74.
Li, Jian; Erte Xiao; Daniel Houser and P Read Montague. 2009. "Neural Responses to Sanction Threats in Two-Party Exchanges." Proceedings of the National Academy of Sciences U S A, 106(39), pp. 16835-40.
List, John. 2004. "Young, Selfish, and Male: Field Evidence on Social Preferences."

Economic Journal, 114, pp. 121-49.
Loewenstein, George. 2000. "Emotions in Economic Theory and Economic Behavior." American Economic Review, 90(2), pp. 426-32.
Lopez, Maria Claudia; James J. Murphy; John M. Spraggon and John K. Stranlund. 2011. "Comparing the Effectiveness of Regulation and Prosocial Emotions to Enhance Cooperation: Experimental Evidence from Fishing Communities in Colombia." Economic Inquiry, doi: 10.1111/j.1465-7295.2010.00344.x, pp. no-no.
Lucas, Robert E. Jr. 1976. "Econometric Policy Evaluation: A Critique," In Carnegie-Rochester Conference Series on Public Policy, 19-46.
MacLeod, W. Bentley. 2007. "Can Contract Theory Explain Social Preferences?" American Economic Review, 97(2), pp. 187-92.
Masclet, David; Charles Noussair; Steven Tucker and Marie-Claire Villeval. 2003. "Monetary and Non-Monetary Punishment in the Voluntary Contributions Mechanism." American Economic Review, 93(1), pp. 366-80.
Meier, B Stefan. 2007a. "A Survey of Economic Theories and Field Evidence on Pro-Social Behavior," In Economics and Psychology: A Promising New Cross-Disciplinary Field, ed. B. S. Frey and A. Stutzer, 51-88. Cambridge: MIT Press.
Meier, Stephan. 2007b. "Do Subsidies Increase Charitable Giving in the Long Run? Matching Donations in a Field Experiment." Journal of The European Economic Association, 5(6), pp. 1203-22.
Mellizo, Philip; Jeffrey Carpenter and Peter Hans Matthews. 2011. "Workplace Democracy in the Lab," In Mimeo.
Mellstrom, Carl and Magnus Johannesson. 2008. "Crowding out in Blood Donation: Was Titmuss Right?" Journal of The European Economic Association, 6(4), pp. 845-63.
Mill, John Stuart. 1867[1848]. Principles of Political Economy with Some of Their Applications. London: Longmass, Green, Reader, and Diver.
Mulder, Laetitia B; Eric van Dijk; David De Cremer and Henk A.M. Wilke. 2006. "Undermining Trust and Cooperation: The Paradox of Sanctioning Systems in Social Dilemmas." Journal of Experimental Social Psychology, 42(147-162).
Murphy, Sheila T.; Jennifer L. Monahan and R. B. Zajonc. 1995. "Additivity of Nonconscious Affect: Combined Effects of Priming and Exposure." Journal of Personality and Social Psychology, 69(4), pp. 589-602.
Murphy, Sheila T. and R. B. Zajonc. 1993. "Affect, Cognition, and Awareness: Affective Priming with Optimal and Suboptimal Stimulus Exposures." Journal of Personality and Social Psychology, 64(5), pp. 723-39.
Rabin, Matthew. 1993. "Incorporating Fairness into Game Theory and Economics." American Economic Review, 83(5), pp. 1281-302.
____. 1998. "Psychology and Economics." Journal of Economic Literature, 36(1), pp. 11-46.
Reeson, Andrew F. and Jhon G. Tisdell. 2008. "Institutions, Motivations and Public Goods: An Experimental Test of Motivational Crowding." Journal of Economic Behavior & Organization., 68(1), pp. 273-81.
Rigdon, Mary. 2009. "Trust and Reciprocity in Incentive Contracting." Journal of Economic Behavior & Organization, 70(1-2), pp. 93-105.
Rodriguez-Sickert, Carlos; Ricardo Andrés Guzmán and Juan Camilo Cárdenas. 2008. "Institutions Influence Preferences: Evidence from a Common Pool Resource Experiment." Journal of Economic Behavior & Organization, 67(1), pp. 215-27.
Rosenthal, Elisabeth. 2008. "Motivated by a Tax, Irish Spurn Plastic Bags," In New York Times. New York.
Ross, Lee and Richard E. Nisbett. 1991. The Person and the Situation: Perspectives of

Social Psychology. Philadelphia: Temple University Press.
Rustagi, Devesh; Stefanie Engel and Michael Kosfeld. 2010. "Conditional Cooperation and Costly Monitoring Explain Success in Forest Commons Management." Science, 330(6006), pp. 961-65.
Rutström, Elisabet 2002. "Sparing the Rod Does Not Spoil the Child: An Experimental Study of Incentive Effects," In Moore School of Business, University of South Carolina.
Schnedler, Wendelin and Radovan Vadovic. 2011. "Legitimacy of Control." Journal of Economics and Management Strategy, Forthcoming.
Schotter, Andrew; Avi Weiss and Inigo Zapater. 1996. "Fairness and Survival in Ultimatum and Dictatorship Games." Journal of Economic Behavior and Organization, 31(1), pp. 37-56.
Seabright, Paul. 2009. "Continuous Preferences and Discontinous Choices: How Altruists Respond to Incentives." The B.E. Journal of Theoretical Economics, 9(Article 14).
Serra, Danila. 2008. "Combining Top-Down and Bottom-up Accountability: Evidence from a Bribery Experiment," In Centre for the Study of African Economies Series. Oxford: University of Oxford. .
Shinada, Mizuhu and Toshio Yamagishi. 2007. "Punishing Free Riders: Direct and Indirect Promotion of Cooperation." Evolution and Human Behavior, 28, pp. 330-39.
Shu, Lisa; Francesca Gino and Max H. Bazerman. 2009. "Dishonest Deed, Clear Conscience: Self-Preservation through Moral Disengagement and Motivated Forgetting."
Sliwka, Dirk. 2007. "Trust as a Signal of a Social Norm and the Hidden Costs of Incentive Schemes." American Economic Review, 97(3), pp. 999-1012.
Sloof, Randolph and Joep Sonnemans. 2011. "The Interaction between Explicit and Relational Incentives: An Experiment." Games and Economic Behavior, In Press, Accepted Manuscript.
Sobel, Joel. 2002. "Can We Trust Social Capital?" Journal of Economic Literature, 40(1), pp. 139-54.
____. 2009. "Generous Actors, Selfish Actions: Markets with Other-Regarding Preferences." International Review of Economics, 56(1), pp. 3-16.
____. 2005. "Interdependent Preferences and Reciprocity." Journal of Economic Literature, 43(2), pp. 392-436.
____. 2010. "Markets and Other-Regarding Preferences," In. San Diego. At: http://federation.ens.fr/ydepot/semin/texte0910/SOB2010MAR.pdf: Department of Economics, University of California. .
Solow, Robert 1971. "Blood and Thunder: Review of the Gift Relationship: From Human Blood to Social Policy by Richard M. Titmuss." The Yale Law Journal, 80(8), pp. 1696-711.
Stanca, Luca; Luigino Bruni and Luca Corazzini. 2009. "Testing Theories of Reciprocity: Do Motivations Matter?" Journal of Economic Behavior & Organization, 71(2), pp. 233-45.
Stiglitz, Joseph. 1987. "The Causes and Consequences of the Dependence of Quality on Price." Journal of Economic Literature, 25(1), pp. 1-48.
Sunstein, Cass R. 1996. "On the Expressive Function of Law." University of Pennsylvania Law Review, 144(5), pp. 2021-53.
Sutter, Matthias; Stefan Haigner and Martin G. Kocher. 2011. "Choosing the Carrot or the Stick? Endogenous Institutional Choice in Social Dilemma Situations." Review of Economic Studies, 77(4), pp. 1540-66.
Taylor, Michael. 1987. The Possibility of Cooperation. New York: Cambridge University Press.
Tenbrunsel, Ann and David M. Messick. 1999. "Sanctioning Systems, Decision Frames

and Cooperation." Administrative Science Quarterly, 44, pp. 684-707.

Tirole, Jean. 1999. "Incomplete Contracts: Where Do We Stand?" Econometrica, 67(4), pp. 741-78.

Titmuss, Richard M. 1971. The Gift Relationship: From Human Blood to Social Policy. New York: Pantheon Books.

Tversky, A. and D. Kahneman. 1981. "The Framing of Decisions and the Psychology of Choice." Science, 211(4481), pp. 453-58.

Tyran, Jean-Robert and Lars Feld. 2006. "Achieving Compliance When Legal Sanctions Are Non-Deterrent." Scandinavian Journal of Economics, 108(1), pp. 135-56.

Vertova, Pietro and Roberto Galbiati. 2010. "How Laws Affect Behavior: Obligations, Incentives and Cooperative Behavior," In SSRN eLibrary. At: http://ssrn.com/paper=1615349.

Warneken, Felix and Michael Tomasello. 2008. "Extrinsic Rewards Undermine Altruistic Tendencies in 20-Month-Olds." Developmental Psychology, 44(6), pp. 1785-88.

Xiao, Erte and Daniel Houser. 2011. "Punish in Public." Journal of Public Economics, 95(7-8), pp. 1006-17.

Young, Peyton and Mary Burke. 2001. "Competition and Custom in Economic Contracts: A Case Study of Illinois Agriculture." American Economic Review, 91(3), pp. 559-73.

Zajonc, Robert B. 1968. "Attitudinal Effects of Mere Exposure." Journal of Personality and Social Psychology Monograph Supplement, 9(2, Part 2), pp. 1-27.

Zhong, C-B; Francesca Gino and Vanessa Bohns. 2010. "Good Lamps Are the Best Police: Darkness Increases Dishonesty and Self-Interested Behavior." Pscyhological Science, 21(3), pp. 311-14.

Zhong, C-B; J Loewenstein and JK Murnighan. 2007. "Speaking the Same Language: The Cooperative Effects of Labeling in the Prisoners' Dilemma." Journal of Conflict Resolution, 51, pp. 431- 56.

Tables 3 to 7.

Note: The bold entries in the comments column -- **I, F, S, E** and **C** -- indicate that the experiment in question could also have been included in tables 3 (**I**nformation about the principal) 4 (**F**raming) 5 (**S**elf-determination) 6 (**E**ndogenous preferences) or 7 (**C**omplementary relations between incentives and social preferences). All the papers but those marked with an * are published or forthcoming in a publication. The entries for each table are organized as follows: First, those studies that are published in a journal, ordered by year and first author. Second, working papers, ordered by year and first author. In a multiperiod experiment, the "stranger" treatment is one in which the pairing of subjects randomly changes from period to period. The "partner" treatment is one in which the pairing of subjects continues from period to period.

Table 3. Bad news: Incentives provide information about the person who implements the incentive (I)

	Citation	Subjects (number)	Games or activities	Institutional environments (treatments)	Results relevant to separability	Comment (quotes are from the cited paper)
[01]	Fehr and Rockenbach (2003)	German students (238)	Trust Game	• Optional punishment as an incentive contract (i.e. a fine if less than the desired back-transfer amount is returned). The level of the fine is fixed by the experimenter and the only choice of the investor is whether to impose the fine or not	Trustee's back-transfers are lower when investors impose fines. Not using the punishment option when it is available results in larger back-transfers and a larger joint surplus.	Explicit incentives undermine altruistic cooperation and reciprocity; forgoing the punishment option is a signal of good will and trust. See Fehr and List (2004). Negative effects of use of the punishment option are greater when the investor demands a larger share of the joint surplus. Categorical crowding out when the investor chooses the fine. **F**
[02]	Fehr and List (2004)	Costa Rican CEOs (126) and students (76)	Trust Game	• Optional punishment as an incentive contract (i.e. a fine if less than the desired back-transfer amount is returned)	CEO principals trust more and are more trustworthy than students and as a result they achieve allocations closer to the maximum surplus that could be generated by the two parties. Joint surplus is highest when the punishment option is available and not used and lowest if the punishment option is used.	Key to performance: "the psychological message…conveyed by incentives – whether ... kind or hostile..." (p. 745). See Fehr and Rockenbach (2003).
[03]	Borges and Irlenbusch (2007)	German Students (179)	Buyer - Seller Game	• Three rights of withdrawal: none, voluntary offer of a right of withdrawal (with a return cost for the seller) and imposed • The right of withdrawal when imposed has a return cost for the buyer or not	When sellers voluntarily offer a withdrawal right, buyers make order decisions that are less harmful for the seller than if the withdrawal right is imposed on sellers exogenously.	"Buyers are more inclined to behave fairly towards the sellers if they have granted the withdrawal right voluntarily than if it is constituted by law". (p. 17) [because it is] "perceived ...as a generous act and they might feel inclined to reciprocate by not exploiting the seller. …". (p. 12). **F**
[04]	Fehr and Schmidt (2007)	German Students (70)	Gift-Exchange Game	• Two internal forms of enforcement: The principal (employer) can choose to rely on - an announced unenforceable bonus contract - a combination of the bonus contract with a fine	Most principals do not use the fine. The joint surplus under the pure bonus contract is 20 percent greater than under the combined contract. Wages are 54 percent higher in the pure bonus contract. Profits are not significantly different in the two contracts.	"Explicit and implicit incentives are substitutes rather than complements" (p. 3). Agents perceive that principals who are less fair are more likely to choose a combined contract and less likely to pay the announced bonus. The effect of effort on the bonus paid is twice as great in the pure bonus case.

Table 3. Bad news: Incentives provide information about the person who implements the incentive (I) (Continued…)

	Citation	Subjects (number)	Games or activities	Institutional environments (treatments)	Results relevant to separability	Comment (quotes are from the cited paper)
[05]	Fehr, et al. (2007)	German students (130)	Gift-Exchange Game	• Three internal forms of enforcement: The principal can choose to rely on - a trust (pure fixed payment) contract, or a price deduction (i.e., fine) contract - a trust, a fine or an unenforceable bonus contract • Different frames: employer-employee or buyer-seller	Under the unenforceable bonus contract subjects contribute more than the payoff maximizing Nash equilibrium, outperforming the enforceable incentive contract (fine). The results are the same independently of the framing.	"Bonus contracts that offer a voluntary and unenforceable bonus for satisfactory performance provide powerful incentives and are superior to explicit incentive contracts when there are some fair-minded players" (p. 121).
[06]	Dickenson and Villeval (2008)	French students (182)	Gift-Exchange Game with a computer task	• Stranger or Partner with communication • Employer payoffs dependent on employee effort (variable) or not	In the partner treatment, when employer payoffs depend on employee effort less monitoring induces substantially higher performance. Consistent with Frey (1993).	While intrinsic motivation is evident in subject behaviors, in the Partner relationship the effect of more monitoring appears to be a reciprocity-based negative response to the principal's lack of trust or intent to benefit at the agent's expense. **F, S**
[07]	Irlenbusch and Ruchala (2008)	German Students (192)	Public Goods Game	• An external form of enforcement (i.e., experimenter imposed): Team-based compensation with and without a reward for the highest contributor in the team • The reward is a low or a high bonus. • Pure individual bonus without team-based compensation	High (but not low) bonuses increase average effort, and joint surplus increases significantly only if the bonus is high, but decreases over time. Only with the purely team-based compensation (no individual incentives) do agents contribute more than self-interest would motivate. Pure tournament incentives induce effort levels below the self-interested Nash equilibrium prediction.	Both categorical and marginal crowding out occur. The tournament structure reduces voluntary cooperation. **F** (See text)
[08]	Ariely, Bracha and Meier (2009)	U.S. students (161)	Charity giving based on task performance	• An external form of enforcement: With monetary compensation or without • Donation choices are public or private • Different frames: "good" and "bad" charitable causes	In the public treatment subjects exert more effort for a good cause and effort is substantially lower in the incentive treatment. Monetary incentives increase effort in the private treatment.	The signaling value of giving is compromised by incentives. "Image motivation is crowded out by monetary incentives [that are] more likely to be counterproductive for public pro-social activities than for private ones." (p.1) Categorical crowding out. Se e Tenbrunsel and Messick (1999), Mulder, et al. (2006).

Table 3. Bad news: Incentives provide information about the person who implements the incentive (I) (Continued…)

	Citation	Subjects (number)	Games or activities	Institutional environments (treatments)	Results relevant to separability	Comment (quotes are from the cited paper)
[09]	Stanca, Bruni and Corazzini (2009)	Italian students (96)	Gift-Exchange Game	• In the first move, Information (player 1 knows there is a second move) or No Information (player 1 does not know there is a second move and hence thinks the game is a Dictator Game)	Second movers' amounts returned are more correlated with the first mover's amounts sent in the No Information treatment.	Reciprocity is stronger in response to actions that are perceived as driven by intrinsic motivation, than to be in response to actions that are perceived as extrinsically motivated. **F**
[10]	Fehr and Gächter (2002b) *	Swiss students (182)	Gift-Exchange Game	• Three external forms of enforcement: A Trust (pure fixed wage) contract, a deduction (i.e., fine) contract, and bonus incentive contract	Incentives reduce agent's effort. If the incentive is framed as a price deduction the effort reduction is greater than where the incentive is framed as a bonus. Incentives reduce total surplus, increase principal's profits.	Effects of incentives are due to the perceived fairness, kindness and hostility of the principal's action. **F, S**

Table 4. Moral disengagement: Incentives may suggest permissible behavior (F)

	Citation	Subjects (number)	Games or activities	Institutional environments (treatments)	Results relevant to separability	Comment (quotes are from the cited paper)
[11]	Hoffman, et al. (1994)	U.S. students (270)	Ultimatum Game; Dictator Game	• Roles are assigned by contest (the right to be the Proposer is 'earned' or randomly assigned). • Different frame: "Exchange" game (between a "seller" and a "buyer") or no frame • Anonymity: Double blind or not	Offers are lower and fewer low offers are rejected in the "Exchange" frame or when the proposer earns the right to his role. Proposers accurately gauge willingness of responders to accept lower offers. Dictators send lower amounts in double blind.	Institutional cues affect behavior: with property rights (i.e. legitimate 'earning' right to be proposer), a market framing or total anonymity proposers and responders are more self-interested. **S**
[12]	Schotter, et al. (1996)	U.S. students (247)	Ultimatum Game;	• Survival treatment (two-stage): subjects with higher payoffs "survive" to proceed to stage 2. • Non survival treatment (one stage): the proposer is randomly assigned • Contextual framing: a simultaneous move-normal or a sequential extensive form game	Competitive threats to survival induce lower offers, and fewer rejections of low offers.	The context affects behavior: 'earning' right to be the first mover or threat to survival induces proposers to behave in a more self-interested manner.
[13]	Cardenas, Stranlund and Willis (2000)	Colombian forest area dwellers (112)	Common Pool Resource Game	• External enforcement device with a low-probability inspection and a fine • Communication	Fines induce more self-interested behavior and are ineffective at reducing common pool overexploitation in the longer run. Socially optimal individual deviations from the self-interested Nash equilibrium behavior (and the implied foregone payoffs by subjects) are least under the fines.	Weakly (exogenously) enforced fines diminish socially motivated behavior. Fine appears to have induced a shift from moral to self-interested frame. See Tenbrunsel and Messick (1999).
[14]	Cardenas (2004)	Colombian users of rural ecosystems (265)	Common Pool Resource Game	• Different levels of external enforcement (low and high fines) with announcement of socially optimal extraction level and without communication • Communication without fines and announcement.	Deviation from self-interested behavior is much greater under communication (no fine) than under either high or low fines without communication. The behavioral effect of high (compared to low) fines is less than 6 percent of the predicted effect assuming self-interested preferences.	Marginal Crowding Out. (See text and also Table 7; where Categorical Crowding In also occurs).

Table 4. Moral disengagement: Incentives may suggest permissible behavior (F) (Continued…)

	Citation	Subjects (number)	Games or activities	Institutional environments (treatments)	Results relevant to separability	Comment (quotes are from the cited paper)
[15]	Heyman and Ariely (2004)	240 US students (150+90)	A computer task and a puzzle task	• Different forms of compensation (cash, candy or a cash-denominated amount of candy) • Different levels of monetary compensation (none, low, medium)	In both the cash and the candy conditions, effort increases when the compensation level increases from low to medium. In the no-compensation treatment, effort is higher than the low-compensation condition for both the cash and the cash in terms of candy conditions and is not different from low-compensation in the candy condition.	The level and form of compensation affect performance. "Monetary compensation may act as a strong signal invoking norms of money markets instead of social-market relations" (p. 6) Monetary incentives influence the ways in which tasks are framed and the motivation to engage in them. The type of market in which the exchange takes place influences the relationship between reward and motivation.
[16]	Bohnet and Baytelman (2007)	Senior executive s in U.S. (353)	Trust Game and a Dictator Game	• No communication, face-to-face pre-play communication or post-play communication • An external form of enforcement (Post-play monetary punishment or not) • Stranger and Partner	Repetition and communication increase amount transferred and back-transferred; the option of punishment for low back-transfers reduces back-transfers of other-regarding trustees (those who send more in the Dictator Game).	"The availability of punishment destroys intrinsic trust and lowers people's willingness to reward trust" (p.1) **I**
[17]	Houser, Xiao, McCabe, et al. (2008)	U.S. students (532)	Trust Game	• A form of enforcement (Punishment as an incentive contract (i.e. a fine)) • Intention treatment: Punishment is assigned exogenously or imposed by investors	When back-transfer requests are high in relation to the sanction's size, regardless of whether the request is fair and regardless of whether punishment is intentional, punishment incentives have detrimental effects on the amount returned.	"Subjects interpret punishment as the price for self-interested behavior and the price, regardless of whether it was intentionally imposed, is an excuse for selfishness" (p.15) Categorical crowding out when the investor chooses the fine. See Fehr and Rockenbach (2003) and Mulder, et al. (2006) **I**

Table 4. Moral disengagement: Incentives may suggest permissible behavior (F) (Continued…)

	Citation	Subjects (number)	Games or activities	Institutional environments (treatments)	Results relevant to separability	Comment (quotes are from the cited paper)
[18]	Mellstrom and Johannesson (2008)	Swedish students (262)	Subjects are offered the opportunity to take a health exam to become blood donors	• With and without a monetary compensation for becoming blood donors • To choose between a monetary compensation and donating the same amount to charity	The incentive reduces the supply of prospective blood donors from 52% to 30% among women. No effect among men. Allowing individuals to donate the payment to charity eliminates the negative effect of the monetary compensation.	The monetary incentive may make it more difficult to signal social preferences, diminishing the signaling value of contributing. Charity option facilitates signaling. Over-justification appears also to be involved. **S**
[19]	Li, et al. (2009)	US citizens (104)	Trust Game	• Optional punishment as an incentive contract (i.e. a monetary sanction if less than the desired back-transfer amount is returned)	Trustees reciprocate relatively less when facing sanction threats, and the presence of sanctions significantly reduces trustee's brain activities involved in social reward valuation (VMPFC, LOFC, and amygdala), while significantly increasing activities in parietal cortex previously implicated in economic decision making.	Monetary sanctions "encourage activity within neural networks associated with self-interested economic decision making while simultaneously mitigating activity in networks implicated in social reward evaluation and processing" (p. 3) **I**
[20]	Henrich, et al. (2010) and Barr, et al. (2009)* and personal communication from Barr and Henrich (March 2009)	15 societies Including US students, African workers, Amazonian, Arctic, and African Hunter-gatherers. (428)	Dictator Game, Ultimatum Game and Third-Party Punishment Game (TPG)	• Differences between societies • Subjects played in the following sequence keeping their role (active or passive): first DG, then the UG and finally the TPG (an explicit incentive, i.e. fine)	In many populations in the TPG the incentives provided by the fine do not induce higher offers, but rather have the opposite effect; factors that may influence self-interest calculations (i.e. wealth, income and household size) are significant predictors of allocations in the TPG (but not in the DG). Membership in a 'world religion' positively associated with offers in the DG but not in the TPG	The presence of the fine in the TPG appears to have reduced the salience of moral reasoning (derived from the teachings of the world religions) and enhanced subjects concerns with their own economic needs. (See text)

Table 5. Control aversion: Incentives may compromise intrinsic motives and self-determination (S)

	Citation	Subjects (number)	Games or activities	Institutional environments (treatments)	Results relevant to separability	Comment (quotes are from the cited paper)
[21]	Gneezy and Rustichini (2000b)	Israeli students (160 for the main experiment)	50 IQ test questions (plus a Principal Agent Game)	• Different levels of monetary rewards for correct IQ test response (very low, low, high and none)	A discontinuity in the effect of incentives at zero. Small rewards degrade performance; large rewards enhance it.	The presence of the incentive substitutes extrinsic for intrinsic motivation. Categorical crowding out. See Gneezy (2003) **F**
[22]	Gneezy and Rustichini (2000b)	Israeli students (180)	Collected donations from households	• Different levels of monetary rewards for the voluntary work (low, high and none)	Discontinuity at zero. Performance with small rewards is lower than performance with high rewards and both are lower than performance with no rewards.	The presence of the incentive substitutes extrinsic for intrinsic motivation. Categorical crowding out. See Gneezy (2003)
[23]	Rutström (2002)	U.S. students (110)	Creative task ('tower of Hanoi')	• Two forms of external enforcement (a penalty or a reward) • Different levels of the external enforcement (none, weak, strong)	Penalties degrade performance; large rewards induce better performance than small (but no better than the no-incentive treatment)	Explicit incentives have a detrimental effect on performance, but only in the case of penalties, not in the case of rewards. Penalties 'distract' subjects. Categorical crowding out.
[24]	Falk and Kosfeld (2006)	Swiss students (804)	Gift-Exchange Game	• Different levels of a lower bound of performance selected by the experimenter (low, medium, and high) • The principal could choose whether to impose the minimum level or not • The principal chooses the agent's wage and whether to impose the bound	Most agents perform minimally (namely at the lower bound) in response to the principals' controlling decision. Majority of the principals anticipate this and do impose the bound, earning higher profits as a result.	Imposing a lower bound compromises subject's sense of autonomy and signals distrust and low expectations that diminish agents' reciprocity and good will towards the principal. Categorical crowding out. (See text) **I**
[25]	Xiao and Houser (2011)	U.S. students (72)	Public Goods Game	• Exogenous punishment: None, private (only the punished subject knows when a round is monitored and the amount of the resulting punishment), public (all members of a group are told that information)	Private punishment induces lower levels of contribution than public punishment.	Weak incentives crowd out cooperation when implemented privately, but the same incentives when implemented publicly (but anonymously) promote cooperation.
[26]	Gneezy (2003)*	US students (400)	Proposer-Responder Game	• The responder has three forms of enforcement (a punishment at a given cost, a reward at a given cost and nothing) • Different levels of the responder's enforcement (weak, strong)	Non-monotonic effects of explicit incentives (fines and rewards) on performance (a W - shaped function). Offers are highest with large incentives (fine and reward), and lowest with small incentives. The no incentive case, when proposers simply dictate allocation, is intermediate.	Extrinsic incentives undermine intrinsic motivation: a small fine or reward changes the mode of behavior from "moral" to "strategic". See Gneezy and Rustichini (2000a), b) and Mulder, et al. (2006). Categorical crowding out. **F**

Table 6. The economy produces people: Incentives alter how new preferences are learned (E)

Citation	Subjects (number)	Games or activities	Institutional environments (treatments)	Results relevant to separability	Comment (quotes are from the cited paper)
[27] Falkinger, et al. (2000) and personal communication from Gächter 18 February 2008.	Swiss students (196)	Public Goods Game	Incentive compatible (Falkinger (1996)) mechanism and no mechanism; large and small group size; Interior and corner Nash equilibria.	Subjects implement the self-interested level of contribution under the mechanism, but contribute substantially more than the self-interested level in its absence (until late in the 20 period experiments) (e.g. Figure 5). After experiencing the mechanism subjects contribute 26 percent less when it is withdrawn than those who have not experienced it.	By rewarding contributions and penalizing shirkers the mechanism may have relieved subjects' sense of moral responsibility and legitimated the pursuit of self-interest. The effects persisted after the withdrawal of the mechanism. **F**
[28] Gneezy and Rustichini (2000a)	Parents from ten day care centers in Haifa, Israel		• An explicit enforcement (i.e. fine) is imposed for lateness in six of these centers.	Tardiness doubles in the six treatment centers and persists even after the fine is removed. No change in the four control centers.	The modest fine may have signaled 'how bad' lateness is and/or is perceived as a price of a service and displaces an ethical frame by a strategic one: "A fine is a price." **I, F, S**
[29] Bohnet, Frey and Huck (2001)	U.S. students (154)	Contract Enforcement Game (finitely repeated)	• Different legal institutions (low, medium or high contract enforcement probability) • Low contract enforcement in the last rounds for all sessions.	The probability of enforcement and/or the cost of breach in the early rounds have a non-monotonic effect on contract performance in the later rounds: intermediate levels of contract enforcement decrease trustworthiness, low levels and high levels of legal contract enforcement increase trustworthiness.	"If there is enough time for the crowding dynamics to unfold, environments with low contract enforcement can produce outcomes as efficient as high levels of enforcement" (p.141) "by affecting behavior, institutions affect preferences." (p.142) **F**
[30] Meier (2007b)	Swiss students (11379)	Contributions to two funds to support financially needy other students.	• Matching donations: For a single semester subjects' contributions are not matched or matched • Matching donations at high or low rates. No matching in subsequent periods	Matching increases contributions when they are in force. But those who experience matching are substantially less likely to make a contribution to either fund in subsequent periods; average contributions show a small, insignificant negative net effect of the incentive.	The negative matching effect is probably not due to the information it conveys on the neediness of the funds (larger effect for the smaller matching rate) or to the subjects' desire to compensate for higher matching induced contributions in the treatment period. **F**

Table 6. The economy produces people: Incentives alter how new preferences are learned (E) (Continued…)

	Citation	Subjects (number)	Games or activities	Institutional environments (treatments)	Results relevant to separability	Comment (quotes are from the cited paper)
[31]	Reeson and Tisdell (2008)	Australian Students (98)	Public Goods Game	• Three external forms of enforcement: - moral suasion in the form of a single sentence to the effect that the payoff to all would be higher if all contributed (all periods); - a binding minimum contribution unexpectedly introduced during 4 periods and then removed - none	While the regulation is in place (during the middle stage) contributions are significantly higher than in the initial stage in which only suasion occurs. After the regulation is removed, contributions are 20 percent lower than in the initial stage. The suasion treatment dramatically increases voluntary contributions compared to a no suasion control.	Suasion enhances and imposed minimum contribution reduces other regarding preferences. Categorical crowding out. **F, S**
[32]	Burks, et al. (2009)	Swiss (139) and US (113) bike messengers	Sequential Prisoners' Dilemma Game	Messenger exposure to performance based pay in their work place or not	In a restricted sample unlikely to be affected by selection bias, second movers' exposure to performance pay is associated with between 12 and 15 percent greater likelihood of defection on a cooperative first mover.	The fact that the effects are from a game having no obvious connection with the job suggests that preferences learned under the incentive conditions of the work place are adopted outside the workplace.
[33]	Irlenbusch and Sliwka (2005)*	German students (84)	Gift-Exchange Game	• Two internal forms of enforcement: The principal can choose - a trust (pure fixed wage) contract - compensation contract (i.e., a variable piece rate) • Two different sequences for the contracts	Incentives reduce cooperation (i.e. effort level) and the effect persists after the incentive is removed. Where principals are constrained to offer fixed wages, the effort levels of agents are considerably higher than when employers can choose an incentive contract.	Incentives (price rate) alter principals' and agents' perception of the situation: "lead agents to adopt an individual maximization frame ... rather than a cooperative frame," "agents have a stronger concern for the principal's wellbeing in the pure fixed wage setting." (p. 23) **F**

Table 6. The economy produces people: Incentives alter how new preferences are learned (E) (Continued…)

	Citation	Subjects (number)	Games or activities	Institutional environments (treatments)	Results relevant to separability	Comment (quotes are from the cited paper)
[34]	Herrmann and Orzen (2008)*	British students (116)	Tullock Rent-Seeking Game or individual choice task and then a Prisoner's Dilemma Game	• Two different sequences (strategic vs. individual): • First stage: the two-player Tullock Rent-Seeking Game (with a different subject) or an individual choice task (with the same incentives). Second stage: a Prisoner's Dilemma	Players cooperate more when they previously played an individual choice task than when the previous game is competitive – strategic, one (i.e. the Rent-seeking Game) Cooperation and reciprocity rates decrease after subjects are exposed to rent-seeking competition.	Subjects may perceive the interaction in the rent-seeking contest as a negative one. "…an individual's attitude towards others undergoes changes between different types of situations because they evoke different contextual cues". (p. 3) "the experience of over-competitiveness in the contest game creates a disposition of rivalry in subjects that some cannot immediately "turn off" when the experiment ends" (p. 26)
[35]	Gächter, Kessler and Konigstein (2011)*	Swiss students (500)	Gift-Exchange Game	• Three external forms of enforcement: a Trust (pure fixed wage contract), a deduction (i.e. fine) contract and a bonus incentive contract • Stranger and Partner • Different sequences	Under incentive contracts agents choose a self-interested best reply (effort) i.e. there is no voluntary cooperation. Experiencing well-designed contracts reduces voluntary cooperation even after incentives are withdrawn.	Incentives may have a lasting negative effect on voluntary cooperation. **F**

Table 7. Incentives crowd in social preferences (C)

	Citation	Subjects (number)	Games or activities	Institutional environment (treatments)	Results relevant to separability	Comment (quotes are from the cited paper)
[36]	Falk, Gächter and Kovacs (1999)	Hungarian students (126, 38)	Gift-Exchange Game	• Stranger and Partner • Two social approval treatments (face to face, social pressure)	Partner treatment increased effort levels; social pressure has little effect. Wage effort relationship (based on reciprocity) is steeper under partner than under stranger.	Repeated interactions provide powerful incentives while enhancing both intrinsic reciprocity motives and concerns for equitable shares (social pressure adds little).
[37]	Gächter and Falk (2002)	Austrian students (116)	Gift-Exchange Game	• Stranger and Partner	With repetition, effort levels are higher than one shot interaction and some self-interested subjects act strategically as reciprocators and then choose the minimal effort level in the last period	Repeated interaction strengthens reciprocity norms and induces 'imitated' reciprocity. "The social norm of reciprocity and the repeated game incentives are complementary." (p.18)
[38]	Masclet, et al. (2003)	US (96) and French (44) students (140)	Public Goods Game	• Two forms of Punishment with different levels of disapproval (from 0 to 10 points received by a subject from any other agent): Monetary punishment (subjects can reduce the monetary payoff of others after observing their decisions) and non-monetary punishment (subjects express disapproval of others' decisions with no effect on others' earnings) • Stranger and Partner • Three stages: In the first and third stages without the punishment. In the second stage, with punishment	Both monetary and non-monetary sanctions induce higher and similar levels of contributions. Individuals tend to make higher contributions relative to the preceding period the higher punishment they have received and the lower their contribution was relative to the group average. When the Punishment device is removed, subjects having been the target of previous monetary sanctions show higher contributions than having non-monetary sanctions.	Cooperation can be enhanced by non-monetary sanctions for reasons that are not strategic and may require repeated interaction. It appears that non-monetary punishment, while not affecting the best response of a payoff maximizing subject, nonetheless raised contributions by enhancing the salience of social motives such as shame or external peer pressure. See Lopez, Murphy, Spraggon, et al. (2011)
[-]*	Cardenas (2004)	Colombian users of rural ecosystems (265)	Common Pool Resource Game	• Different levels of external enforcement (low and high fines) with announcement of socially optimal extraction level and without communication • Communication without fines and announcement.	Deviation from the self-interested Nash extraction level was 29% greater under the small fine than with no fine.	Individuals consider the norm of cooperation that is proposed externally [the announced optimal level] when extracting (p. 238). Categorical crowding In. (See text and also Table 3; where marginal crowding out also occurs)

* This reference is not numbered since it is an additional result of an already cited study [14].

Table 7. Incentives crowd in social preferences (C) (Continued…)

	Citation	Subjects (number)	Games or activities	Institutional environment (treatments)	Results relevant to separability	Comment (quotes are from the cited paper)
[39]	Henrich, et al. (2005)	Foragers, herders, others in 15 small-scale societies (1128)	Ultimatum Game	• Differences between societies in the level of market integration and the potential payoffs to cooperation	Substantial cross cultural co-variation between the degree of market integration (engagement in market exchange) and both average UG offers and (unpublished) the propensity to reject low offers.	Mutually beneficial interactions in market interactions with strangers may support the evolution of cultures of fair-mindedness towards strangers; "*doux commerce*"? Hirschman (1977). This study also presents evidence of incentives alter how new preferences are learned **E**
[40]	Falk, Fehr and Zehnder (2006)	Swiss Students (240)	Labor Market Game (one employer, three workers)	• With and without a minimum wage. • Two different sequences	The introduction of a legal minimum wage affects workers' fairness preferences leading to a rise in their reservation wages (which persists even after the minimum wage has been removed).	"Minimum wages [may] affect [subjects'] fairness perceptions" (p.1376) creating moral "entitlements". Obligations activate and or enhance social preferences. See Galbiati and Vertova (2008), Vertova and Galbiati (2010)
[41]	Tyran and Feld (2006)	Swiss students (102)	Public Goods Game	• Levels of sanctions: none, mild and severe • Enforcement: external or self-imposed (by referendum)	Experimenter imposed mild sanctions do not significantly affect average contributions to the public good. Compliance is much improved if mild law is endogenously chosen.	Experimenter imposed sanctions raised the expected cost of freeriding without affecting behavior; a possible explanation is that only referendum imposed sanctions conveyed a signal of moral disapproval by peers.
[42]	Herrmann, et al. (2008a)	16 student pools around the world (1120)	Public Goods Game (Partner)	• Monetary Costly Punishment	Cooperation is higher in the punishment condition. However, the average payoff with the punishment condition is lower than the average without punishment in many countries. Weak norms of civic cooperation and the weakness of the rule of law in a country are significant predictors of antisocial punishment (targeting high contributors), which reduces the net benefits to the group.	Punishment is socially beneficial only if complemented by strong social norms of cooperation with strangers so that peer punishment induces shame rather than resentment. The quality of the formal law enforcement institutions and informal sanctions are complements.
[43]	Rodriguez-Sickert, Guzmán and Cárdenas (2008)	Rural Colombians from 5 communities (128)	Common Pool Resource Game	• Three different forms of external enforcement (A fine regime imposed, a fine proposed to the players and rejected or accepted by them, none) • Different levels of external enforcement (low, and high) for the imposed fine	Under all treatments other than no fine, groups start at high levels of cooperation. Cooperation remains high only when a fine, be it high or low, is in force. If the players reject the fine, cooperation slowly unravels. Presence of low fines prevented unraveling of cooperation.	When fines are rejected, the implied affirmation of social norms may have temporarily increased cooperation; reciprocal preferences (anger at low contributors) may account for the subsequent erosion of cooperation. Small fines enhance unconditional cooperation by relieving cooperators of the need to retaliate against defectors by withdrawing their own cooperation.

Table 7. Incentives crowd in social preferences (C) (Continued…)

	Citation	Subjects (number)	Games or activities	Institutional environment (treatments)	Results relevant to separability	Comment (quotes are from the cited paper)
[44]	Carpenter, et al. (2009)	US students (172)	Public Goods Game	• Costly punishment: subjects can punish non-cooperators at a cost to themselves • Different team's residual claim (marginal per capita return on the public good) • Different group size	Shirkers are punished by peers and they respond by contributing more, even in the last round unless the frequency of reciprocators is too low or the group is too large. High contributors who are punished subsequently contribute less. (Unpublished results not reported in paper).	Altruistically motivated mutual monitoring, by enhancing shame-induced cooperation, supports high levels of team performance. Synergistic effects of social preferences and peer-imposed incentives. This study also presents evidence of incentives alter how new preferences are learned **E**
[45]	Carpenter and Myers (2010)	U.S. Volunteer firefighters (217) and non-volunteer community members (189)	Dictator Game	• Exogenous variation in the presence and level of small stipends paid to volunteer firefighters	Small monetary incentives increase turnout to fighting fires for firemen unconcerned about image but have no effect on image-concerned firemen (the estimated negative effect is not significant).	The effect of image concerns increases with the visibility of the activity (training is a less visible activity than fighting fires). For firefighters with image concerns the positive direct effect of small extrinsic incentives is offset by the negative indirect effect of incentives on the image-value of fighting fires.
[46]	Gächter, Nosenzo and Sefton (2010)	British Students (84)	Gift-Exchange Game with 3-member firms (one employer and two employees)	• Employees move sequentially (Employee 1 has pay comparison information (i.e. information about what coworker earns) and Employee 2 additionally has effort comparison information (information about how co-worker performs) • Employers can offer high wages to both employees, a high wage to Employee 1 only, a high wage to Employee 2 only and low wages to both	A homogeneous wage does not affect effort when an employee is matched with a co-worker that provides less effort. Reciprocity toward the employer is more pronounced when the co-worker is hard-working, as effort is strongly and positively related to own wage and when the employer pays unequal wages to the employees. Exposure to pay comparison information in isolation from effort comparison information does not appear to affect reciprocity toward employers.	Unequal wages conditional on worker type may induce high levels of reciprocity based effort; unconditional employer generosity fails to recognize the 'deserving' worker, and is not reciprocated. Incentives and social preferences as complements. Workers respond to employers' recognition of their work effort and hence deservingness, not to employer generosity.

Table 7. Incentives crowd in social preferences (C) (Continued…)

Citation	Subjects (number)	Games or activities	Institutional environment (treatments)	Results relevant to separability	Comment (quotes are from the cited paper)
[47] Lopez, et al. (2011)	Colombian Fishermen (180)	Public Goods Game	• Public reminder about benefits of cooperation plus 1/5 chance of receiving private reminder of the social losses resulting from the individual's non-cooperative behavior (Guilt); receiving public reminder of the social losses resulting from the individual's non-cooperative behavior (Shame), facing an external low penalty for not contributing to the public good, facing an external high penalty for not contributing.	Priming subjects to feel guilty about low contributions did not affect average contributions, but the random public revelation of one's contributions (inducing shame) substantially increased contributions. Experimenter's imposition of the fine further increased contributions but the level of the fine has no effect.	Results suggest the importance of moral framing and that the fine did not work as an incentive but rather as a signal highlighting the salience of the ethical dimension of the problem. Categorical Crowding in.
[48] Vertova and Galbiati (2010)*	Italian students (210)	Public Goods Game (and a Lottery Game)	• Different levels of a stated (non-binding) obligation to contribute (zero, low and high) • A symmetric incentive structure (a level of contribution less (more) than the minimum contribution could be subject to a probabilistic penalty (reward)) with low and medium size	There is a positive effect of the obligation, which is greater when it is combined with a weak monetary incentive than when no incentives are offered. A stronger monetary incentive does not result in an increase in contributions. The strong monetary incentive also has no effect on behavior in the absence of the stated obligation.	Incentives not only influence material payoffs but also frame recommended high contributions as obligations. Incentives and obligations affect people's behaviors by activating values and/or coordinating individuals' beliefs. See also Galbiati and Vertova (2008)
[49] Barr (2001)*	Zimbabwean villagers (602)	Public Goods Game	• Two external forms of non-monetary punishment i)Public announcement: each player announces her level of contribution to everyone present in the session, ii) Subjects could make public verbal statements about each other's decisions: lighthearted criticism or the withholding of praise during informal gatherings	After the introduction of the public announcement and public criticism subjects contribute more.	The fact that non-material punishment raises contributions suggests that it induces shame or other social emotions (the best response for a self-interested individual was unaffected). See Gächter and Fehr (1999) and Mulder, et al. (2006).
[50] Serra (2008)*	British students (180)	Bribery Game (public official-citizen)	• Three different forms of external enforcement (no monitoring; top-down auditing, and an accountability system which gives citizens the opportunity to report corrupt officials)	Under the accountability system, fewer officials engage in corruption. The presence of only top-down auditing did not affect the number of officers who demanded a bribe but induced corrupt officials to demand a higher bribe than no monitoring.	"Non-monetary costs activated by the bottom-up component of the combined system had a significant impact on the public official's decision to engage in bribery." (p.17)

II. Pro-social behavior, Heterogeneity and Incentives: Experimental evidence from the local commons in Colombia

Abstract[3]

Promotion of pro-social behavior in social dilemmas has been subject to an increased interest among social scientists and policy makers due to the relevance of Common Pool Resources (CPR) in long-run human well-being. Although economists know a lot about the effect of incentives on pro-social behavior, we understand very little about how to promote pro-social behavior efficiently. For example, we accept that monetary incentives sometimes are less effective than would be predicted for entirely self-regarding individuals but we are unable to identify the mechanisms by which these crowding effects may occur.

This study examines a unique experimental data set of a CPR game with 1095 individuals (79% are CPR users that are closely related to a real resource). Our purpose is twofold. On one hand, it goes ahead on the experimental analysis by accounting for unobserved heterogeneity of individuals' social preferences and group composition of types within each group. On the other hand, it explores the role of heterogeneity of pro-social preferences on achieving the most efficient economic incentive.

First, we drop the assumption that all individuals are all self-regarding and develop several models of pure Nash strategies for our CPR game when individuals are motivated by a combination of self-interest and preferences for altruism, reciprocity or inequity aversion. Second, we estimate individual heterogeneity by using a random coefficients model approach and classify individual social preferences (according to their behavior in the baseline phase) by assigning a type to every participant. Third, we compare the role of heterogeneity of preferences in social efficiency across incentives and confirm the existence of different effects of incentives on each type; the subsidy is found to be the most socially efficient incentive. Finally, we obtain *exogenous* determinants of individual type such as level of education, perceptions on the CPR, perceived interest in cooperation among the community, whether the participant does volunteer work and whether the CPR is the household main economic activity of the household; we also obtain *endogenous* determinants such as the composition of types in the group and their demographic characteristics.

JEL classification: A13 (Relation of Economics to Social Values), C23 (Random effects models), C51 (Model Construction and Estimation), D64 (Altruism), H39 (Fiscal Policies and Behavior of Economic Agents), Q20 (Renewable Resources and Conservation), Q28 (Renewable Resources and Conservation, Government Policy).

Keywords: Common-pool resources, social preferences, laboratory and field experiments, explicit incentives, inequity aversion, altruism, random coefficients model

[3] I thank Syngjoo Choi for detailed comments and suggestions. I am grateful to David Echeverry, Sam Bowles, Orazio Attanasio, Steffen Huck, Juan-Camilo Cárdenas, Ana Roldán and Adriana Molina. I also acknowledge financial support to the fieldwork from the MacArthur Foundation and the Javeriana University in Colombia.

1. Introduction

Promotion of pro-social behavior in social dilemmas has been subject to an increased interest among social scientists and policy makers due to the relevance of Common Pool Resources (henceforth CPR) in long-run human well-being (Janssen et al. (2010)). The motivation for this study is that researchers in economics know a lot about the existence of pro-social behavior. An individual behaves pro-socially in order to help others -including himself, to achieve a common good. Social preferences are those concerns for the well-being of others and desires to uphold ethical norms. They reduce social inefficiency in the absence of complete contracts (Arrow (1971); Becker (1976); Akerlof (1984)) and thus are the key to solve social dilemmas (Ostrom (1990)), in which the uncoordinated actions of individuals result in an outcome that is Pareto inefficient[4].

Although economists know a lot about the effect of incentives on pro-social behavior, we understand very little about how to promote pro-social behavior efficiently. Most of Mechanism Design and Public Economics models assume that individuals are self-regarding, that is, they face own-material-payoff-maximizing motives. Although it has been found that economic incentives[5] may be less effective than predicted (Sobel (2005); Bowles (2008)). We remain unable to identify the mechanisms by which these crowding effects may occur[6].

It is also acknowledged that allowing for heterogeneity is key to achieve more efficiency in public policies (Kreps (1997)). This approach implies that the design and evaluation of public policies must consider the composition of social preference types in the target group and the effect of incentives on each type. However, how incentives interact with heterogeneous social preferences remains mostly unexplored; existing experimental evidence and economic theory do not complement each other when trying to solve this puzzle. This study presents evidence on the role of heterogeneity in social preferences on efficiency of incentives.

The aim of the paper is twofold. First, introduces a structural approach to examine which types of social preferences individuals exhibit in a CPR environment, in which the CPR is collectively owned or shared among populations (e.g. natural resources, land, software) and foregoing the overexploitation of the jointly used resource leads to a Pareto superior outcome. Second, it examines the effect of heterogeneity of social preferences in achieving social efficiency of incentives. This is especially important for policy makers who look for not only effective but also efficient incentives. Understanding heterogeneity of individual preferences in this environment is the first step to design Pareto efficient incentives.

We address these questions, first by using a CPR experimental game in the field to find empirical evidence of heterogeneity of social preferences types and explaining these types of social preferences within an economic model. Second, we use this structural model to have a robust estimation of the theoretical parameters and the distribution of types in our sample.

---

[4] The individual optimization problem is to maximize i's payoff, i.e. $\max_{x_i} \pi_i \left(x_i, \sum_j x_j\right)$. Given that $\frac{\partial \pi_i}{\partial x_i} > 0$ and $\frac{\partial \pi_i}{\partial \sum_j x_j} < 0$, the Pareto optimal allocation does not coincide with the individually rational strategy.

[5] Which I will call simply "incentives" without the adjectives, meaning interventions to influence behavior by altering the economic costs or benefits of some targeted activity.

[6] See Bowles and Polania-Reyes (2012) for a review and Holmes et al. (2010) for empirical attempts.

Second, we assess the impact of incentives on social efficiency via their effect on the different types of behavior. We find that individual pro-social behavior is consistently different across types and the composition of types in the group significantly determines the initial average level of extraction for the group, the path of extraction and the outcomes. Besides, different incentives have a different effect on behavior, which is determined, in part, by the composition of types in the group.

Finally, we examine the possible determinants of being classified as a certain type by taking into account socio-economic characteristics and the composition of types in the group.

The paper goes as follows: next session highlights this paper's contribution. Section 3 introduces the CPR framework and the experimental design for this study. We also compare the models' predictions with and without an incentive (i.e. low subsidy, low fine, medium fine, high fine and very high fine). In Section 4, we drop the assumption that all individuals are all self-regarding and develop several models of pure Nash strategies for our CPR game when individuals are motivated by a combination of self-interest and preferences for altruism (Levine (1998)), reciprocity (Bowles and Gintis (2003)) or inequity aversion (Fehr and Schmidt (1999)). Section 5 describes the data and our experimental procedures. In section 6, we classify individual social preferences according to their behavior in the baseline (no incentive) phase. We estimate individual heterogeneity by using a random coefficients model and examine incentives' performance in the light of efficiency and efficacy. We also compare the role of different preferences in the social efficiency of incentives, and confirm the existence of different effects of incentives on each type and the best efficient incentive. In Section 7 we relate the probability of being type q and the theoretical parameters of altruism and reciprocity to both *exogenous* variables (such as socio-economic characteristics, perceptions on the CPR and the community and social capital measures) and *endogenous* ones such as the composition of the group in terms of types and observable characteristics. The last section concludes and suggests points of future research.

2. Heterogeneity of social preferences

While there is a vast literature that show a substantial fraction of the players are motivated by motives different from self-interest, there has been little evidence on the coexistence of different social preferences present among individuals. Theoretical and empirical advances over the intervening years provide the basis for a reconsideration of the simultaneous presence of different social preferences (Rabin (1998); Kreps (1997); Elster (1998), Leider et al. (2009)) and the effectiveness of incentives to promote pro-social behavior not only in the lab, but also in the field (Bowles and Polania-Reyes (2012)). On the other side, the role of economic incentives and pro-social behavior has been studied mostly in Principal-Agent settings (Kessler (2008); Falk et al. (2005); Falk and Kosfeld (2006)) with few exceptions (see for example Rodriguez-Sickert et al. (2008)).

This paper contributes to provide additional implications for policy design and mechanism design theory. First, we drop the assumption that when individuals exhibit social preferences, they also show the same type (Hwang and Bowles (2014); Dufwenberg et al. (2011); Charness and Rabin (2002)). Second, we improve the type classification method, which is the most used currently in the literature of social preferences (i.e. random coefficients model) and

explain these types of social preferences within an economic model[7]. Finally, we assess the role of heterogeneous social preferences on the efficiency of incentives.

We contribute to the literature of Common Pool Resource games that have been implemented by assuming homogeneous preferences mostly with only students (Falk et al. (2002)) based on (Walker et al. (1990)) and only real users (Margreiter et al. (2005); Velez et al. (2009)). Our sample is unique since which has both students and real users of the CPR and different CPR environments (water, firewood and fish) under a rich set of (economic and non-economic) incentive schemes. This dataset has been used before to show that users deviate more from self-regarding NE than students (Cárdenas (2004, 2011) and Cárdenas and Carpenter (2005)) and the effect of incentives at the average level (Cardenas et al (2004) and Cardenas (2005)). In addition, among real users of the CPR, those whose main resource of income is extraction from the CPR deviate more from self-regarding NE (Molina (2011)). All these studies explain these findings with the existence of social preference but do not explore the role of heterogeneity of social preference among the participants.

This paper is closely related to Velez et al. (2009). Our aim too, is to understand better the behavior of small-scale common pool resource users in developing countries. We also include different models of social preferences. However, we merge preferences for fairness and reciprocity as in Ledyard (1995) and Rabin (1993) but introducing a social norm measured as participants' initial belief in the practice rounds. Another difference with respect to Velez, et al.'s specification is that that we look at consistency of behavior to categorize an individual instead of characterizing aggregate behavior. Finally, we examine the role of incentives in social preferences and the effect of having different types in a group on efficiency.

Finally, this paper contributes to the literature on structural models of pro-social behavior (Falk and Heckman (2009); Manski (2011)) since we use the random coefficient model approach to estimate theoretical parameters. One of our strengths is that we have more than one observation per individual, which allows us to examine certain types of consistent behavior among individuals, contrary to other studies that use one-shot decision-classification. On the other hand, our results suggest the need to estimate a distribution for the –unobservable- categorization of social preference types and the parameters of such preferences, which is possible thanks to new advances in econometrics such as finite mixtures models (Echeverry and Polania-Reyes (2015)).

3. Common Pool Resource framework

i. Static Common Pool Resource game

This model was first presented in Cárdenas (2004) and Cárdenas et al. (2004). First, we are going to define the game $\Gamma = \{X^n, f\}$. An individual i is endowed with e units of effort (i.e. hours of extraction, investment in equipment, etc.) which he can use to extract $x_i \in X = \{1, \dots, e\}$ units from the CPR.

[7] In applied microeconomics, social preferences types have been studied in the context of public goods games based on Walker et al (1988) and risk preferences (see Andersen, et al. (2010)) and predominantly with only students (Fischbacher et al. (2001); Kurzban and Houser (2001); Burlando and Guala (2005); Bardsley and Moffat (2007); Carpenter et al. (2009)) or with only real users (see evidence in forest management by Rustagi et al., (2010)). For a good commentary, see Velez, et al. (2009) and Rustagi et al., (2010).

(1) $$f: \quad X^n \to \Pi^n$$
$$(x_1, \dots, x_n) \mapsto (\pi_1, \dots, \pi_n)$$

The payoff function π_i is defined as

(2) $$\pi_i = ax_i - \frac{1}{2}bx_i^2 + \varphi\big(ne - (x_i + x_{-i})\big)$$

whose components are the direct benefits from extraction $ax_i - \frac{1}{2}bx_i^2$ and the indirect costs from depletion $\varphi\big(ne - (x_i + x_{-i})\big)$ by i's decision and other's decision, x_{-i} of extraction.

In this particular case, we assign to the fixed parameter the values $(a, b, \varphi, e) = (60, 5, 20, 8)$ and the number of players in the group g is $n = 5$, thus, they face a payoff function $\pi_i = 60x_i - \frac{5}{2}x_i^2 + 800 - (x_i + x_{-i})$, the set of possible choices of individual i are $X = \{1, \dots, 8\}$ and the possible aggregate level of extraction is given by $(x_i + x_{-i}) \in \{5, \dots, 40\}$.

The unique socially (Pareto) efficient allocation or social optimum (SO), which also maximizes the aggregate payoffs of the group $\sum_i \pi_i$ s.t. $x_i \in X$ and it is given by $(x_1^{SO}, \dots, x_n^{SO}) = argmax_{(x_1,\dots,x_n)} \sum_{i=1}^{n} \pi_i$. The optimization conditions yield the optimal decision of extraction $\hat{x}_i^{SO} = \frac{a-\varphi n}{b} < 0$ and our feasible socially optimal decision of extraction is a corner solution $x_i^{SO} = 1$ of extracting the minimum level possible of the resource[8].

The Social Efficiency Index (SEI) according to the number of individuals in the group g is defined as the ratio between the actual aggregate payoff, or the sum of the individual payoffs, $\sum_i \pi_i$ and the maximum aggregate payoff possible, the sum of the individual payoffs when all the individuals in the group play the social optimal strategy, $\pi_i^{SO} = \pi\big(x_i^{SO}\big) = 758 \; \forall i$. Under the social optimum, all players extract one unit and $SEI_g = \frac{\sum_i^5 \pi_i}{\sum_i^5 \pi_i^{SO}} = 1$.

There is a conflict between the self-regarding Nash equilibrium and the socially efficient strategies (See table A1 in the appendix). As we explain in detail below, the Unique Nash Equilibrium (SNE) for a self-regarding individual is given by $x_i^{SNE} = argmax_{x_i} \pi_i \; \forall i$ that is $x_i^{SNE} = \frac{a-\varphi}{b} = e = 8$, given the chosen parameters (a, b, φ, e). If there are only self-regarding individuals the social efficiency index is $SEI_g^S(x^{SNE}) = 0.42$.

ii. Dynamics of the CPR game: experimental design

Participants played a finitely repeated partner matching game, ($T = 10$ periods with $t = \{1, \dots, T\}$) with common knowledge of the payoff function. At the beginning of period t each individual decides simultaneously her level of extraction, x_{it}. At the end of period t, the experimenter announces aggregate extraction$(x_{it} + x_{-it})$ and players are informed about

[8] Since x_{it} takes only non-negative values, for framing the experimental design we have a corner solution at $\hat{x}_i^{SO} = 0$, that is players should not allocate labor into extraction to produce the socially efficient outcome. We have eliminated in the payoffs table that option to avoid possible conflicts in conducting these experiments in the field. Previous experiments and pre-testing exercises in the field suggest that there is a strong aversion towards prohibition of resource use that could create problems with the participants when conducting the experiments this way. Interior solutions with non-dominant strategies, such as used in Ostrom, Gardner and Walker (1994) and Cardenas et al. (2000), are another alternative, but here we also have decided to maintain corner solutions in order to have a design with a dominant strategy (Cárdenas (2005):250).

other players' aggregate behavior. That is i does not know individual extraction by the other players, x_{-it}. She only knows the average extraction by them, $\bar{x}_{-it} = \frac{\sum_{j \neq i}^{n-1} x_{jt}}{n-1}$.

Before round 11 the experimenter announces an incentive is implemented (see Figure A1 and the Appendix section for instructions and procedures). The groups remain unchanged during the following T rounds $t = \{11, \dots, 20\}$ (i.e. partner matching). The incentive could be a non-monetary incentive (i.e. the incentive does not affect individual payoffs) or a monetary incentive (fine or subsidy).

Some sessions did not face any incentive as we needed control groups, that is, during the remaining ten rounds individuals had the same baseline instructions.

iii. Monetary incentives to promote prosocial behaviour

In order to prevent over exploitation of the CPR, an economic incentive s is introduced, which is proportional to the level of extraction[9]. For each round, one player is chosen randomly with an exogenous probability of inspection $p = \{0.1, 0.2\}$ and face one monetary incentive. When the monetary incentive is a fine, s is the monetary cost of violating the rule of the socially optimum solution. If violating the rule, that is if has a level of effort that is greater than 1, $x_i > x^{SO} = 1$, the player pays a fine $s < 0$ and $s \in \{-350, -175, -100, -50\}$ per unit of extraction above 1. In particular, equation (3) now includes the new expected payoff function for a player i, which is as follows:

$$\pi_i^{s<0} = \pi_i + ps(x_i - x^{SO}) \tag{3}$$

The new Self-regarding Nash Equilibrium is $x_i^{SNE\_s<0} = \frac{a-\varphi+ps}{b} = \{1,6\}$ for $x_i \in [1,8]$. When the monetary incentive is a subsidy, s is the monetary reward for extracting less than the Self-regarding Nash strategy. If the chosen player's level of extraction such a level, $x_i < x^{SNE} = 8$, the player receives a subsidy, $s = 50$ per unit of extraction below 8. The expected payoff function for a player i, is now as follows:

$$\pi_i^{s>0} = \pi_i + ps(x^{SNE} - x_i) \tag{4}$$

The new Self-regarding Nash Equilibrium is $x_i^{SNE\_s>0} = \frac{a-\varphi-ps}{b} = 6$, for $x_i \in [1,8]$. The Social Efficiency Index, SEI, for a self-regarding individual is 62% for $s \in \{-100, -50, 50\}$ and 100% for $s \in \{-350, -175\}$. Table 1 summarizes the different monetary incentive schemes.

Table 1. Incentive scheme and new Self-regarding NE for each incentive

Incentive s (addition to π_i)		*Acronym*	**Individual level of extraction x_i** 1	2	…	7	8	$x_i^{SNE\_s}$	**Probability of inspection p**
Subsidy $+s(8-x_i)$	**Low**	**SL**	350	300	. . .	**50**	0	6	0.2
Fine $-s(x_i-1)$	**Low**	**FL**	0	**-50**	. . .	-300	-350		0.2
	Medium	**FM**	0	**-100**	. . .	-600	-700		0.1
	High	**FH**	0	**-175**	. . .	-1,050	-1,225	1	0.2
	Very high	**FVH**	0	**-350**	. . .	-2,100	-2,450		0.1

[9] The data set also contains additional treatments in which the monetary incentive is not exogenous. Players could decide whether to implement a monetary incentive such as i) and ii) at round 11 voting on whether to implement a low fine (or subsidy) and players decide who should be inspected, iii) and iv) a low fine (or subsidy) at each round, v) voting for a low subsidy at each round, and vi) and vii) voting a high (or low) fine at round 11.

iv. Non-monetary incentives to promote prosocial behavior

In order to prevent over exploitation of the CPR, individuals face one exogenous non-monetary incentive. These are *one-shot communication* (OC) in which the group has a single five min face-to-face communication only once prior to making all ten decisions (corresponding to rounds 11 to 20); *repeated communication* (RC) in which the group has a single five min conversation before each round; or *public announcement* (PA), in which the rule of extracting the social optimum level is announced. For the last treatment, in each round, one player is chosen randomly with $p = 0.2$; if violating the rule ($x^{SO} = 1$), the player pays no fine but must show the experimenter his or her extraction level, which is then announced publicly to the group. None of these non-monetary incentives alter the self-regarding Nash prediction (SNE).

v. Adaptive dynamics with heterogeneous types

Note that the one-shot static model doesn't correspond to the experimental design with T periods and we have to introduce time dynamics to the model in the estimation. Thus, it is necessary to set up the primitives of the adaptive dynamics of the individuals. For ease of computation we assume individuals have adaptive expectations[10]. Each type has a belief about others' decision x_{-i} in period t

(5) $$B_t(x_{-it}) = \bar{x}^e_{-it} = \bar{x}_{-it-1} \ \forall i$$

where $\bar{x}_{-it-1} = \frac{\sum_{j\neq i}^{n-1} x_{jt-1}}{n-1}$ is other players' average level of extraction in the previous period. Although we are aware this is an extreme assumption, it is still useful for this study since we are focus on the classification procedures. The dynamics of the process of belief formation will be considered in further research. In addition, the initial condition $\bar{x}_{-i1} = \bar{x}_{-i0}$ we assume this one to be equal to the average of the individual level of extraction in the practice round; which is equal to 4.51 in our data.

4. Models of individual social preferences

In this section we present a structural approximation of individual prosocial behavior and consider the most popular types of individuals in the behavioral economics literature: Self-regarding, altruist, reciprocator and inequity averse. A risk neutral individual i maximizes a utility function U_i defined as[11]

(6) $$X^n \xrightarrow{g} \Pi^n \xrightarrow{U_i} \mathbb{R}_+$$
$$(x_1, \dots, x_n) \mapsto (\pi_1, \dots, \pi_n) \mapsto U_i(\pi_1, \dots, \pi_n)$$

[10] This is relevant to the analysis of repeated games and relies on reinforcement or other forms of real-time learning. See Rabin (1993), Dufwenberg and Kirchsteiger (2004), and Falk and Fischbacher (2006).

[11] For simplicity, we are assuming that the individual utility function is linear in the players' payoffs. For example, in the case of a self-regarding individual, $U_i^S = U_i(\pi_i) = \varrho\pi_i$ with $\varrho = 1$. However, neutrality is an important matter measuring social preferences. It is necessary to perform the same analysis with other functional forms of the utility function. For example, when $U_i(\pi_i) = \frac{(\pi_i + c_i)^{1-\varrho}}{1-\varrho}$ where c_i is individual consumption of goods and $U_i(\pi_i) = \frac{(\pi_i + bw_i)^{1-\varrho}}{1-\varrho}$ where w_i is individual wealth and $b > 0$. This issue can be explored with the data set we have and is part of our current research.

The decision process for each individual is to maximize the expected individual utility function:

$$\max_{x_i} \mathrm{E}_{B_t} U(\pi_{it}, \bar{\pi}_{-it-1}; \Theta) \tag{7}$$

where $\bar{\pi}_{-i} = \frac{\sum_{j \neq i}^{n-1} \pi_j}{n-1}$ is other players' average payoff and $\Theta = \{\rho_i, \mu_i, \beta_i\}$ is a vector of individual parameters for the individual utility function, $U(.)$. We will now describe different models of social preferences and the maximizing strategies for every type q of social preferences from their respective utility function, U_i^q.

i. Baseline: self-regarding preferences

Individuals that exhibit self-regarding preferences care only about their own monetary cost and benefits and are usually called in the literature as free riders, selfish and defectors. A *Self-regarding individual* i has a utility function given by $U_i^S = \pi_i$. The Self-regarding best-reply (from the first order conditions) is $x_i^S = \frac{a-\varphi}{b} = 8$ and the Self-regarding Nash equilibrium is given by the maximum individual level of extraction or $x_i^{SNE} = 8$ units in our CPR framework $\forall i$.

Result 1. *Individuals that play the self-regarding NE strategy are a small proportion of the sample.*

This result is not new in the literature of social preferences. The first generation of experimental evidence reveals the existence of social preferences in which people may consistently deviate from the self-regarding Nash prediction and that social preferences are important influences on economic behavior (Fehr et al. (1997); Bewley (1999); Fehr and Schmidt (1999); Fehr and Gächter (2000)). When looking at individual behavior over rounds we find that the first result to motivate this study is that individuals that play the self-regarding NE strategy are a small proportion of the sample. Figure 1 presents the distribution of times that individuals extract 8 units among the first ten rounds. There are 35 percent of the players who never played the NE strategy, a quarter of the players chose the NE strategy only once and we observe that only 2 percent of 1000 individuals played the NE consistently during all periods.

We have to consider that there are identification issues between reputation and learning. For example, individuals may behave pro-socially in the presence of reputation effects (Kreps et al. (1982); Bohnet and Huck (2004); Mailath and Samuelson (2006)) (see Figure B1 in the appendix). We affirm this is not the case since the players have complete information on the payoffs and know that the game is finitely repeated. Thus, the sub-perfect SNE of this finite game is the same for all the rounds. This is consistent with models with incomplete information about other's preferences and/or that allow for reputation in games (Mailath and Samuelson (2006))[12]. In addition, although the fraction of players that extract 8 units is small at the beginning and increases at the last period, it never reaches 25% of the sample.

We also observe that individuals deviate from the NE strategy most of the time. We have a glimpse on heterogeneous behavior in our sample since we are called to explore what the 70% of the sample did choose. Then we need to go to a higher level of cognition in order to explain this behavior and use our definition in section 3. This approach and the individual

[12] This is also confirmed by the assumption that there is no Social multiplier in the game.

preferences framework are key in order to answer our research question. Now we will present an individual preferences framework in a static setting for different types of social preferences.

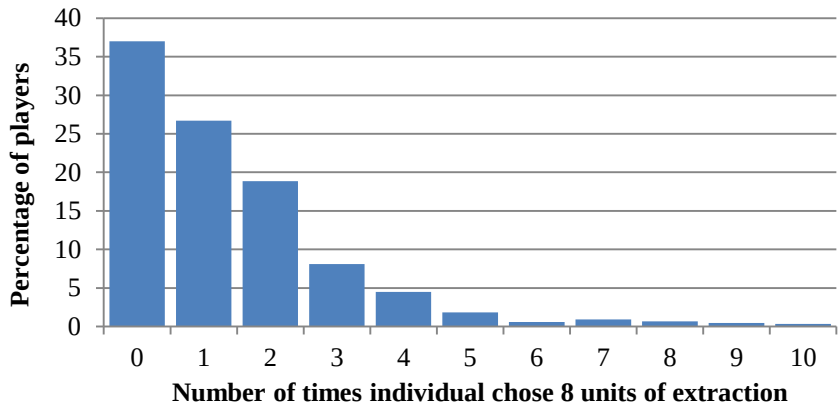


Figure 1. Percentage of individuals that behave as Self-regarding. N=1095 (CPR users and students). In the x-axis we observe the number of extractions $x_i = 8$ for during the first ten rounds. In total, only 0.35% chose the NE strategy consistently during the first ten rounds.

ii. Altruistic preferences

We adapt our CPR framework to the models proposed by Levine (1998) and Casari and Plott (2003). Individuals that exhibit these preferences are those who care about other's utility (such as altruism (Andreoni and Miller (2002); Carpenter et al. (2009)), unconditional cooperator (Fischbacher et al. (2001)) or pure cooperators (Rabin (1993)). An Altruist has a utility given by

(8) $$U_i^A = \pi_i + \rho_i \bar{\pi}_{-i} \text{ with } \rho_i \in (0,4]\ \forall i$$

where ρ is the parameter of altruism, the positive weight an altruist puts on other's payoff. Note that this is equivalent to $\rho_i \in (0,1]$ if $\pi_{-i}$is used instead of $\bar{\pi}_{-i}$. The altruistic best reply (from the first order conditions) is[13]

(9) $$x_i^A = \begin{cases} x_i^S - \frac{\rho_i \varphi}{b} = 8 - 4\rho_i & if\ 0 < \rho_i < 1.75 \\ 1 & if\ 1.75 \leq \rho_i \leq 4 \end{cases}$$

If there are only pure cooperators, that is individuals with a very high ρ, we will have $\rho_i \geq 1.75\ \forall i$ and we obtain the Pareto efficient outcome, $x^{SO} = 1$.

iii. Reciprocal preferences and a social norm

Reciprocators cooperate only if others cooperate and present a similar behavior to conformism (Rabin (1993); Bowles (2004); Levine (1998)). When individuals do not have complete information about other's behavior, they use the current social norms which stem from beliefs about others' behavior. A social norm is a pattern of behavior such that individuals prefer to conform to it on the condition that they believe that most people in their reference network i) conform to it (i.e. empirical expectations) and ii) think they ought conform to the norm (i.e. normative expectations) (Bicchieri (2006, 2014)) . Given that decisions are private and individual in the CPR game, the game is able to capture empirical expectations the first time they play (i.e. the practice rounds). Empirical expectations are key

[13] These are boundary conditions given $X = \{1, \ldots, 8\}$: $0 \leq \rho_i/4 \leq 0.44$.

for social norms to evolve and they are mostly based on observations of what individuals in the reference group have done in the past (Bicchieri (2014)). In addition, in repeated encounters, people have an opportunity to learn from each other's behavior, and to secure a pattern of reciprocity that minimizes the likelihood of misperception (Bicchieri and Moldoon, (2014)). We then define x_i^* is an ethical prescription governing actions towards others, a social norm determined by culture, reference points or the context of individual behavior. In order to examine reciprocity, we use as social norm $x_i^* = \overline{x_0} = 4.51\ \forall i$, the average number of extracted units in the last practice round among CPR users in the first visit since at that stage subjects had not formed their expectations on which types they were interacting with in their group.

A nonaltruistic *Reciprocator* individual i (exhibits neither good will nor spite unconditionally but conditions her behavior on the goodness or spitefulness of others) has a utility given by

(10) $$U_i^R = \pi_i + \mu_i(x_i^* - \bar{x}_{-i}^e)\bar{\pi}_{-i}\ \forall i$$

where the weight the reciprocator puts others' payoffs, μ_i, is positive and based on which i rates extractions from others deriving more utility if others' expected average extraction, $\bar{x}_{-i}^e$, is below this norm and less otherwise. The reciprocal best reply (from the first order conditions) is given by

(11) $$x_i^R = \begin{cases} 8 & if\ x_i^* \leq \bar{x}_{-i}^e \\ 8 - 4\mu_i(x_i^* - \bar{x}_{-i}^e) & if\ x_i^* > \bar{x}_{-i}^e \text{ and } \mu_i \in \left[0, min\left(1, \frac{1.75}{x_i^* - \bar{x}_{-i}^e}\right)\right] \end{cases}$$

An individual with preferences for *altruism* and *reciprocity* (Bowles, 2004) has a best reply function given by

(12) $$x_i^{AR} = \delta_{1i} + \delta_{2i}\bar{x}_{-i}^e\ \forall i$$

where $\delta_{1i} = 8 - 4\mu_i x_i^* - 4\rho_i$ and $\delta_{2i} = 4\mu_i$. In the empirical section we are going to estimate the structural parameters ρ and μ by using the Random Coefficients model.

iv. Preferences for fairness and inequity aversion

The model of fairness and inequity aversion is based on Fehr and Schmidt (1999) and Bolton and Ockenfels (2000). We use the adaptation to a CPR model by Falk, et al. (2002). An *inequity averse* individual i has a utility given by

(13) $$U_i^I = \pi_i - \alpha_i \max(\bar{\pi}_{-i} - \pi_i, 0) + \beta_i \max(\pi_i - \bar{\pi}_{-i}, 0)\ \forall i$$

The second term in Equation (13) measures the utility loss from disadvantageous inequality, and the third term measures the loss from advantageous inequality. It is assumed that the utility gain from i's payoff is higher than her utility loss for advantageous inequality and her utility loss from disadvantageous inequality is larger than the utility loss if player i is better off than other players, $0 \leq \beta_i < 1$. In addition, i is loss averse in social comparisons: i suffers more from inequality that is to his disadvantage (Loewenstein et al. (1989)), $\alpha_i \geq \beta_i$[14].

There are two cases to study when comparing individual payoffs. First, if $\pi_i \leq \bar{\pi}_{-i}$ there

[14] However, given our parameters, this is never the case in our game. Since the Nash equilibrium is a corner solution, there is no possibility of disadvantageous inequality.

is *disadvantageous inequality*. The critical condition is that $x_i^- \leq \frac{a(1+\alpha_i)-\varphi}{b(1+\alpha_i)} \leq 8$ that yields a parameter $\alpha_i = 0\ \forall i$. This is common when the Self-regarding NE is the upper bound of the feasible set: in this case when there is disadvantageous inequality, i will want to extract more units up to the maximum possible.

The second case is when there is *advantageous inequality* or $\pi_i > \bar{\pi}_{-i}$. The critical condition is $x_i^+ \geq \frac{a(1+\beta_i)-\varphi}{b(1+\beta_i)} \geq 1$ which yields a set of symmetric equilibria with individual's best reply (from the first order conditions)

$$(14) \qquad x_i^I = \begin{cases} (8 - \frac{4\min\beta_i}{1-\min\beta_i}, 8] & if\ 0 \leq \beta_i < 0.63 \\ 1 & if\ 0.63 \leq \beta_i < 1 \end{cases}$$

and $\min\beta_i$ is the smallest β_i among all n players in the group. Table 2 presents a summary of this section. In the next section we estimate the structural parameters for an individual that may exhibit altruistic or reciprocal preferences.

Table 2. Defining Θ: Individual's utility functions and best responses in a static setting

Individual type label $U_i^q(\pi_i, \bar{\pi}_{-i}; \theta_i^q)$	i's Best response	Restriction	Parameter $\Theta = \{\theta_i^q\}$
Self-regarding $U_i^S = \pi_i$	$x_i^S = 8$		$\{0\}$
Altruist $U_i^A = \pi_i + \rho_i \bar{\pi}_{-i}$	$x_i^A \in [8 - 4\rho_i, 8)$ $= 1$	$0 < \rho_i < 1.75$ $1.75 \leq \rho_i \leq 4$	$\{\rho_i\}$
Reciprocator $U_i^R = \pi_i + \mu_i(x_i^* - \bar{x}_{-i}^e)\bar{\pi}_{-i}$	$x_i^R = \mathrm{h}(\mu_i, x_i^*, \bar{x}_{-i}^e)$ $= 8$	$x_i^* > \bar{x}_{-i}^e$ $x_i^* \leq \bar{x}_{-i}^e$	$\{\mu_i\}$
Inequity averse $U_i^I = \pi_i - \alpha_i \max(\bar{\pi}_{-i} - \pi_i, 0) + \beta_i \max(\pi_i - \bar{\pi}_{-i}, 0)$	$x_i^I \in (8 - \frac{4\min\beta_i}{1-\min\beta_i}, 8]$ $= 1$	$0 \leq \beta_i < 0.63$ $0.63 \leq \beta_i < 1$ and $\alpha_i = 0$	$\{\beta_i\}$ only if $\pi_i > \bar{\pi}_{-i}$

Note: We have chosen a functional form $U_i^q(\pi_i, \bar{\pi}_{-i}; \theta_i^q) = \pi_i + \theta_i^q \bar{\pi}_{-i}$ with $\bar{\pi}_{-i} = \frac{\sum_{j \neq i}^{n-1} \pi_j}{n-1}$. In our experimental setting the n-player game thus becomes a **2-player** game, though each of the n = 5 players is playing a different game. Given $\pi_i = 60x_i - \frac{5}{2}x_i^2 + 800 - (x_i + x_{-i})$, the individual's best response does not depend on $\bar{x}_{-i}^e = \frac{\sum_{j \neq i}^{n-1} x_j^e}{n-1}$, except for reciprocators. This will help for identification purposes in the analysis. For an example, see Appendix C.

5. The Data and experimental design

The experiments were conducted in 8 Colombian rural villages (see Table B1 in the appendix) between 2000 and 2002 and a university in Bogotá. A total of 1095 participants attended the sessions, 230 undergraduate students and 865 real users of a CPR. Participants in the sessions in the field were not only familiar with the use of common-pool resources – such as fisheries, water or firewood – but also knew each other and had a prior history of reputation building before the experimental sessions. Every village depended on a different CPR[15].

In a session of twenty rounds, each of five players in a group has to decide a level of extraction between 1 and 8 units of a resource during two stages of ten rounds each. The game was framed as a one situation in which individual households have to decide about the

[15] This would make the use of type-detection and reading of intentions by players during the experiment more salient (Cárdenas et al., 2004).

extraction level of a resource such as fish, firewood or water. In each round, a monitor collected decisions and recorded them privately and confidentially. The monitor added the individual extraction levels and announced the total extraction for the group in that round. By knowing the group extraction and their own individual extraction, players were asked to calculate their individual earnings according to the payoff table (see Appendix A). However, players did not know the individual decisions of the others in the group, just their aggregate extraction. This procedure was repeated for twenty rounds. At the end of all the rounds, earnings were added for each participant, and each was privately paid in cash[16].

Table 3. Real Users' Socio-economic Characteristics (N=865)

Variable		Mean	Median	Min.	Max.	Standard deviation	Obs.
Household Size		5.02	5	1	19	2.75	773
Age		34.03	32	7	85	13.91	765
=1 if Woman		46.86	0	0	1	0.50	764
Years of education		5.98	5	0	18	3.73	703
=1 if Landowner		75.03	1	0	1	0.43	749
Number of children		2.28	2	0	6	1.93	359
=1 if membership in associations		44.42	0	0	1	0.50	824
=1 if attends at least to one meeting		62.01	1	0	1	0.49	824
Number of associations meetings		19.82	8.5	1	720	56.20	360
=1 if individual has done any volunteer work in the past year		65.58	1	0	1	0.48	738
Number of days of volunteer work per year		18.91	1	0	365	62.86	738
Percentage Households whose one of two main economic activities is	Agriculture	40.17	0	0	1	0.49	824
	Cattle	11.77	0	0	1	0.32	824
	Fishing	35.32	0	0	1	0.48	824
	Wood	6.01	0	0	1	0.24	815
	Hunting	2.21	0	0	1	0.15	815
Fraction of households with CPR extraction as main economic activity	100%	7.04					824
	50%	31.31					
	0%	61.65					
Time spent by households on CPR extraction	100%	6.19					
	50%	31.07					
	0%	62.74					

All data were collected using standard procedures in experimental economics in the laboratory: no deception, no field referents, and fully salient choices. See Cárdenas (2004) and Cárdenas et al., (2004) for the details on this experimental design and its modeling, Cárdenas et al. (2000) for the first field experiment with an external regulation tested experimentally in the field, and Cárdenas (2011) for experimental procedures on the field

[16] Each experimental session in the field was conducted with five people who lived in the same village. Groups were randomly formed, but we avoided members of the same household participating in the same session. All sessions in the field and the university were hand-run, using pencil and paper. Each session took about three hours. This included the time taken for reading the protocols (see Appendix A), running two or three practice rounds, undertaking two stages of 20 rounds, and calculating the earnings by subject while they filled out a short exit survey and other demographic details (for the villagers only). During the first stage of ten rounds, all sessions were run under the same set of rules. The subjects were notified that the experiment would last at least ten rounds, and that during these rounds no communication among themselves would be allowed. The villagers were seated in a circle facing outwards so that the privacy of the decisions could be maintained. Once the ten rounds were over, the monitor announced that a second stage was about to start, under a new set of rules, for another ten rounds. None of the groups knew in advance during the first ten rounds the type of new rules for the second stage of the game (for more details on the experimental procedures see Appendix A and Cárdenas, et al. (2004) and Cárdenas (2005)).

experiment data set used in this paper[17]. We collected information on individual characteristics of CPR users only. Table 3 summarizes the individual characteristics of CPR users that participated in the games.

On average the household size was 6 people, the average age of the participants was 34 years old and nearly half of the participants were women. We also collected information on the CPR, which are shown in the appendix. To our knowledge this is the richest dataset on CPR in a developing country[18]. Table 4 summarizes the sample allocated according to the treatments and phases of an experimental session.

Table 4. Sample and treatments

Phase	Treatment	Individuals	Groups		
			Total	CPR users	Students
Round 1-10	Control	1095	219	173	46
Round 11-20	Low Subsidy	105	21	18	3
	Low Fine	125	25	20	5
	Medium Fine	50	10	5	5
	High Fine	80	16	12	4
	Very High Fine	25	5	3	2
	Total Pure Economic Incentive	*385*	*77*	*58*	*19*
	Communication one shot	80	16	13	3
	Communication each round	75	15	13	2
	No FINE- public announcement	60	12	9	3
	Total Non-Economic Incentive	*215*	*43*	*35*	*8*

6. Identification of social preferences types: the Random Coefficients model approach

The Random Coefficients model (RCM) is a classification method based on Swamy (1970); Kurzban and Houser (2005); Carpenter and Seki (2011) and Carpenter et al., (2009). A backward looking individual i who has a maximum level of extraction of $e = 8$ units and decides her level of extraction $x_i \in [1,8]$ may have a characteristic behavioral attribute. The parameter heterogeneity is treated as stochastic variation In order to obtain individual measures of social preferences exhibited in the game, we use a random coefficient estimation, which assumes the coefficient vector is the outcome of a random process and the parameters are drawn from a population of possible coefficients and allows the estimates to vary by individuals. The output of the model is an intercept and slope for each subject.

i. Estimation

[17] Individuals did not know how many rounds they would play. There were two example rounds and one practice round and the game started once the experimenter ensured the participants understood the procedure.
[18] There is a caveat to take into account when analyzing CPR users' decisions. In 2002, three of the villages that participated in 2001 were visited for a second time. Individuals that participated in the experiments in this second visit, i.e. 160, although not all had participated before showed a significantly different behavior compared to the overall sample of individuals who participated before, i.e. 705. We did not want to exclude the sample of the second visit since it contains also valuable information about group behavior. For now on, we will describe in the text individual behavior of the overall sample unless we state otherwise. All the information and comparison between the first and second visits is examined in Cárdenas and Carpenter (2005). See Cárdenas et al. (2015) for a novel proposal to explain heterogeneous behavior with this sample.

In the baseline (i.e. no treatment) the level of extraction of individual i depends on the average of the level of extraction by the other members in the group in the previous round $(t-1)$ such that:

$$x_{it} = \delta_{1i} + \delta_{2i}\bar{x}_{-it-1} + \epsilon_{it} \quad \forall t = 1,..10 \tag{15}$$

where individual behavior has two dimensions: δ_{1i} is a measure of Altruism or unconditional cooperation, δ_{2i} is a measure of Reciprocity or conditional cooperation, ϵ_{it} is iid $N(0, \sigma^2_{\epsilon_i})$ and the coefficient vector for the ith individual such that $\delta_i = \delta + \nu_i$ with $\nu_i \sim N(0, \Sigma_\nu)$[19].

The estimation of the structural parameters of reciprocity and altruism is given by equation (15), $\hat{\rho}_i = 2 - \hat{\mu}_i x_i^* - \hat{\delta}_{1i}/4$ and $\hat{\mu}_i = \hat{\delta}_{2i}/4$.

ii. Classification of types

There are four types of social preferences in this model and each one must satisfy different conditions, shown in Figure 2 and summarized in Table 5. First, a Self-regarding individual presents a high level of 'unconditional' extraction $\hat{\delta}_1 \geq 6$ for all the rounds $t = 1,..,10$ and a low coefficient of reciprocity $|\hat{\delta}_2| \leq 1/4$. Second, a reciprocator or "If others extract less, I will extract less"- individual exhibits a positive slope $\hat{\delta}_2 > 1/4$. Third, a "*Counter*" or "If others extract less, I will extract more" individual exhibits a negative $\hat{\delta}_2 < -1/4$[20]. Finally, an Altruist individual presents a low level of 'unconditional' extraction $\hat{\delta}_1 \leq 3$ for all rounds $t = 1,..,10$ and a low coefficient of reciprocity $|\hat{\delta}_2| \leq 1/4$.

Table 5. Type classification conditions using the RCM approach

	Social preferences' type q			
Parameters' conditions	Altruist	Self-regarding	Reciprocator	Counter
Unconditional extraction, $\hat{\delta}_1$	$\hat{\delta}_1 \leq 3$	$\hat{\delta}_1 \geq 6$	$\hat{\delta}_1 \leq 6$	$\hat{\delta}_1 \geq 3$
Reciprocity, $\hat{\delta}_2$	$\lvert\hat{\delta}_2\rvert \leq 0.25$	$\lvert\hat{\delta}_2\rvert \leq 0.25$	$\hat{\delta}_2 > 0.25$	$\hat{\delta}_2 < -0.25$

Our best algorithm that maximizes the number of individuals allocated to a certain type does not consider test of significance since the number observations per individual is ten[21]. Note that altruists are also non-reciprocators. It is possible to find reciprocators whose decision was to extract less but conditioned on whether others extracted less. Figure 2 presents the zones where the individuals should be allocated given this classification method procedure. Right in the middle of the space, there are those individuals that do not meet any of the cases, called the '*unidentified*' individuals.

[19] In addition, $(\delta_1, \delta_2)_i = (\delta_1, \delta_2) + \nu_i, E(\nu_i) = 0, E(\nu_i \nu_i') = \Sigma$.

[20] Duffy and Munoz-Garcia (2013) show the existence of a 'backstabber' player type who is unconcerned about fairness initially cooperates in order to disguise himself as a player type who is concerned about fairness and inducing the uninformed player to cooperate in all periods of the repeated game until the 'backstabber' takes the opportunity to defect, i.e., he "backstabs" the uninformed player.

[21] However, when using a restriction of level of significance for $\hat{\delta}_2$ significantly different from zero at 90%, the proportion of individuals allocated as a reciprocator and a counter falls 33.19% for first time - CPR users and 30.86% for students.

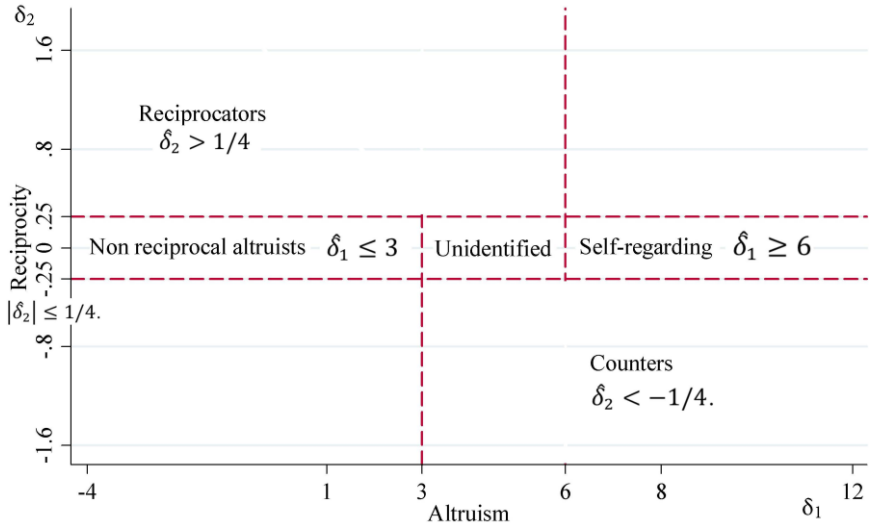


Figure 2. Heterogeneity with the RCM

Result 2. *Using the RCM classification method, we find there is heterogeneity of social preferences.*

Figure 3 shows the distribution of observed heterogeneity among CPR users. In our classification we did not consider the complete sample since students and CPR users show significantly different behaviors (See Appendix, Figure D1 and Figure D2). Every dot is an individual lying in the $(\hat{\delta}_1, \hat{\delta}_2)$ space. Among the first visit-CPR users, 29.22 percent (229 individuals) where found to be the reciprocator type, 10.35 percent are the altruistic type, 6.95 percent the self-regarding type and 21 percent the counter type. Although the number of counters seems large, we observe their behavior as a common one; when some 'super' free riders or even spiteful, wait to build trust among other players in order to extract more and obtain a higher payoff in one round (see figure B1a in the appendix). Unfortunately, this specification approach does not allow us to include an advantageous inequality averse type (i.e. the β parameter)[22]. Individuals that could not be categorized are called type "unidentified" are 32.5 percent of the individuals.

Table 6 shows the estimates for a random coefficient regression model for CPR users in the first visit[23]. The values and signs are the expected. Individuals that are considered the altruist type have in average a $\hat{\rho}_i$= 1.40 whereas those considered self-regarding would have face a $\hat{\rho}_i$= 0.37. Note that this value is not zero because Self-regarding individuals are those who would consistently extract 6 or more units. On the other hand, reciprocators have the highest value of $\hat{\mu}_i$, 0.13 and those we call Counter have the opposite value, -0.13.

[22] When implementing a model that allows for preferences for reciprocity, altruism and inequity aversion, we do not obtain enough exclusion restrictions in order to identify β_i. However, we conducted an RCM classification estimation by using only the cases when $\pi_i > \bar{\pi}_{-i}$. Our results (available upon request) are similar to the model with reciprocity, confirming evidence by Falk et al (2002) which demonstrate that self-interest combined with inequity aversion generates a best-response function with the same form as the self-interest and reciprocity best response function.

[23] For the same table with students and CPR users in the second visit see table D1 and D2 in the appendix. The coefficient size is substantially larger for CPR users than for the students, which shows a much stronger shift towards cooperative choices for the former.

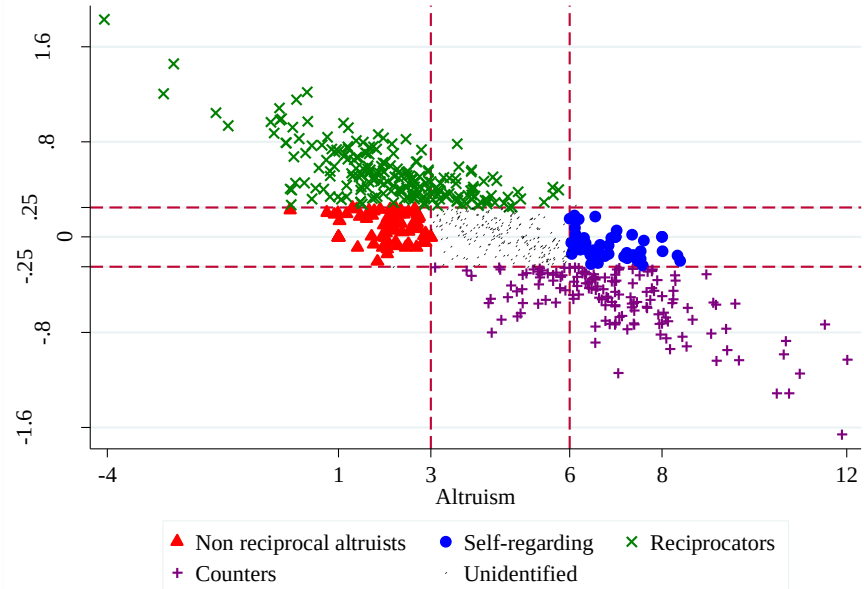


Figure 3. The distribution of observed heterogeneity of social preferences in the game with the RCM – First time visit - CPR users (N=705)

Table 6. Individual structural parameters estimators of Altruism and Reciprocity for each type – First visit CPR users

Structural parameter		Non reciprocal Altruists	Reciprocator	Self-regarding	Counter	Unidentified
Altruism, $\hat{\rho}_i$	Mean	1.40	0.86	0.37	0.85	0.84
	Standard Dev.	0.19	0.27	0.17	0.32	0.20
	Minimum	1.07	0.11	0.00	-0.06	0.16
	Maximum	1.77	1.70	0.65	1.83	1.73
Reciprocity, $\hat{\mu}_i$	Mean	0.02	0.13	-0.02	-0.13	0.00
	Standard Dev.	0.03	0.06	0.03	0.06	0.03
	Minimum	-0.05	0.06	-0.06	-0.41	-0.06
	Maximum	0.06	0.46	0.05	-0.06	0.07
N		73	206	49	148	229

Notes: For CPR users in the first visit, we set $x_i^* = 4.51$, the average number of extracted units in the last practice round.

iii. Examining efficiency

The RCM procedure gives us '*rough*' social preference parameters, given the assumptions we have to make. However, we are able to validate this modeling approach by showing with a sensitive analysis how our assigned types behave in terms of efficiency in their groups. Consistently, among students, the average SEI was 0.67 whereas among CPR users in the first visit the average social efficiency index was 0.73[24]. Now, we examine how the type of neighborhood affects social efficiency.

Result 3. *Social preferences composition affects efficiency.*

This effect is different from what we would expect by simple aggregation. In Table 7, we observe the Social Efficiency Index SEI_g^q according to the number of type q individuals in the group g and the fraction of individuals type q in the group of $n = 5$ players[25]. In the first column there is the number of individuals in the group of 5 players that were categorized by

[24] Among CPR users in the second visit the average social efficiency index was 0.78.
[25] For this table with the student and 2nd visit CPR users samples, see Tables D3 and D4 in the appendix.

the RCM of type q. A dot denotes absence of observations when there was no group with a certain composition. When there is a dot For example, we may have groups with no altruist and all reciprocators. Among the 141 groups of CPR users in the first visit, there were never 4 altruists in a group and there were no groups with 4 or 5 self-regarding individuals. When the fraction of altruists predominates in the group the SEI yields the highest (92.27 percent) value over all groups. We compared efficiency between groups with the same number of individuals being type q by doing a t-test of means (Altruist vs. Self-regarding, self-regarding vs. positive reciprocator, positive vs. negative reciprocator, negative reciprocator vs. counter and counter vs. unidentified). Column two in Table 6 shows that when the number of altruists increases in the group, social efficiency is significantly higher than efficiency when the number of self-regarding individuals increases, that is, when there are many predicted self-regarding in the group, the average payoff and social efficiency in the group are the lowest. The fifth column presents a decreasing SEI from 69.3 to 63.14 when the number of self-regarding individuals in the group increases up to 3 people. Another highlight is that when there were no altruists in the group efficiency was comparable to having mostly self-regarding in the group.

As we defined above, we use as social norm $x_i^* = \overline{x_0} = 4.51\ \forall i$, the average number of extracted units in the last practice round among CPR users in the first visit. The second and third columns of table 6 show that the reciprocators respond to the social norm: efficiency decreases with the number of reciprocators in the group where extraction is higher than the norm and increasing where extraction is lower than the norm. Among reciprocators, the SEI reaches to the maximum when there are three reciprocal individuals and it is higher when the average level of extraction by others is lower than the benchmark, x_i^*. The initial level of extraction matters in order to respond positively when this level is high or negatively when it is low according to the amount of reciprocators in the group.

Finally, although the 'counter' type is a high fraction of the sample, having a high number of 'counters' in the group is not significantly different for those who could not be categorized, i.e. unidentified[26].

Table 7. Social efficiency and group heterogeneity with the RCM- CPR users first visit

No. individuals in group g that are type q	Individual Type q label					
	Altruist	Reciprocator		Self-regarding	Counter	Unidentified
		$x_i^* > \bar{x}_{-i}^e$	$x_i^* \leq \bar{x}_{-i}^e$			
0	69.81***	75.44	68.96	75.02	73.20	74.90
1	76.35**	76.51	69.59*	69.30**	72.92	74.58
2	83.81***	78.79	68.25**	64.87**	73.83	72.61
3	85.80	80.37	65.48**	63.14	71.25	69.78
4	.	68.78	63.26	.	73.56	70.56
5	92.27	77.15	69.05	.	.	71.78

Unpaired t-test of difference in means between consecutive types (i.e. altruist vs. self-regarding, reciprocator when $x_i^* > \bar{x}_{-i}^e$ vs. reciprocator when $x_i^* \leq \bar{x}_{-i}^e$, reciprocator when $x_i^* > \bar{x}_{-i}^e$ vs. self-regarding, reciprocator when $x_i^* \leq \bar{x}_{-i}^e$ vs. counter and counter vs. unidentified) having the same number of players type q. ***, ** and * respectively represent significance at 1%, 5% and 10% confidence level. The number of groups is 141.

Result 4. *Incentives have heterogeneous effects on social efficiency across groups. Social preferences composition affects incentives performance.*

[26] Although we would like to examine whether it matters who else is in the group, it is not possible to compare the SEI across the different type composition and vary the type the majority is playing with since we lack of enough observations.

Incentives affect individual and group behavior differently from what is predicted by the model of Self-regarding individuals. We look at group's efficiency according to the group's composition of types and the growth rate of efficiency for each incentive[27]. First, we performed the same analysis in table 6 in the presence of every incentive. For example, in Table 8a we observe how social efficiency for each group composition was affected by communication. Among the different types the highest impact of the incentive is when most individuals in the group were reciprocators (65.48 vs. 88.84).

Table 8a. Social efficiency and group heterogeneity in the presence of communication at each round- CPR users first time visit

No. individuals in group g that are type q	Individual Type q label					
	Altruist	Reciprocator		Self-regarding	Counter	Unidentified
		$x_i^* > \bar{x}_{-i}^e$	$x_i^* \leq \bar{x}_{-i}^e$			
0	85.39	.	.	89.07	89.68	90.78
1	93.18	89.28	78.20	86.26	88.99	87.27
2	93.58	88.70	87.75	.	86.95	92.30
3	.	91.07	88.84	.	.	83.19

In Table 8b we observe how a low fine affected social efficiency for each group composition. The performance of the incentive is very poor, since there are small improvements in efficiency. Among the different types the highest impact of the incentive is when there was one self-regarding individual in the group (89) and when there were no altruist in the group (85.38) and the lowest when most individuals in the group were reciprocators (65.48 vs. 88.84).

In addition, under the High fine (not shown), negative reciprocators, i.e. $x_i^* \leq \bar{x}_{-i}^e$ respond prosocially with an increase of 30% in efficiency. Self-regarding individuals cooperate more only with the low subsidy: when there are three (and two) self-regarding individuals with an increase of 37% (and 39%) increase in efficiency. The lower growth rates of efficiency are those by positive reciprocators, i.e. $x_i^* > \bar{x}_{-i}^e$ under the very high fine with a decrease of 15% in efficiency and altruists with an increase of 8% in efficiency under a fine. Overall negative reciprocators' behavior improves.

Table 8b. Social efficiency and group heterogeneity in the presence of a low fine- CPR users first visit

No. individuals in group g that are type q	Individual Type q label					
	Altruist	Reciprocator		Self-regarding	Counter	Unidentified
		$x_i^* > \bar{x}_{-i}^e$	$x_i^* \leq \bar{x}_{-i}^e$			
0	85.38	82.73	72.95	85.91	83.33	87.19
1	77.89	84.67	73.86	89.00	88.72	88.96
2	97.25	89.73	82.57	75.58	82.35	85.48
3	.	83.56	68.34	.	.	72.25

We calculate the growth rate in efficiency for each different group composition and define the SEI growth rate as the ratio $SEI_{g\ t>10}^{q\ s}/SEI_{g\ t\leq 10}^{q} - 1$ for every incentive s, number of individuals q in group g. In Table 8c we observe the average of SEI growth rates by incentive across groups; thus observing efficiency by taking account heterogeneity. The control group presents a natural average decrease by 0.88%, a result consistent with the worsening of cooperation across time when there is no regulation. All groups are better off

[27] In order to obtain a comparable measure of efficiency for the low subsidy, we extracted from the earnings the prize the inspected individual obtained by deviating from the SNE.

with communication every round. Efficiency rates grew up to 35.67% and 20.29% on average.

Result 5. *Non-economic incentives perform better than economic incentives. Communication increases efficiency.*

Among the monetary incentives, the low subsidy had the best performance, and the high fine is the second best, with a efficiency growth rate up to 41% and 15.89% on average, despite of the decrease in efficiency among altruist with a -1%. Efficiency reaches the highest efficiency among self-regarding with the low subsidy with an average increase of 19.38%. The least effective incentive is the very high fine.

Table 8c. Incentives and Social Efficiency Growth rate across groups

Treatments	Efficiency Growth rate		
	Average	Maximum	Minimum
Control	-0.88	6.64	-10.18
Low Subsidy	*19.38*	*39.25*	*4.81*
Low Fine	13.30	28.43	2.02
Medium Fine	11.38	24.38	-4.61
High Fine	15.89	40.99	-1.03
Very High Fine	3.20	25.03	-19.29
Total Fine Incentive	*13.32*	*27.20*	*0.02*
Communication one shot	14.68	39.91	0.83
Communication each round	20.29	35.67	11.65
Total Non-monetary incentive	*17.92*	*39.91*	*5.00*

The following three institutions evaluate the impact of imposing a sanction on non-compliance with a probability of 1/5 of being inspected. The heavier sanction comes with the high fine, which shows an impact similar to that of the communication in one shot. On average, the fine as an incentive increases efficiency 13.32% lower that the improvement by non-monetary incentives 17.92%. As the empirical experimental evidence on CPR games suggests (Ostrom et al., 1994, Hackett et al. 1994, Ahn et al., 2011, Zhosan and Gardner, 2013), the most efficient incentive is communication each round with an average increase of 20.29%.

iv. State-dependent social preferences

We proceed to implement the random coefficient model estimation for the second stage of the experiment, rounds 11 to 20 and compared the differences in type classification according to the incentive.

Result 6. *Exposure to treatment affect individual's type*

In table 9a and Table 9b we present the effect of the incentive on individual behavior by types for CPR users in the first visit. In the second column we show the classification generated with the first ten rounds. From column 3 to 7, we show for each type assigned in the baseline the distribution of the individuals among different types in the treatment phase. Table 9a shows the shifts in social preferences after the implementation of communication one shot and communication before each round. For both treatments, self-regarding individuals disappeared. Altruist did not change their behavior when communication occurred at every round but when communication occurred only once, 44% of altruist became reciprocators and 11% were unidentified. The highest percentage of individuals who

changed their type were those of counter: 50% and 36% became altruist in the second stage. These findings are consistent with the hypothesis that communication helps players to detect the types of players the hypothesis that communication helps players to detect the types of players with whom they are interaction or the actions that others are more likely to take (Frank (1988), Brosig (2002) and Cárdenas et al., (2004))[28]

Table 9a. Percentage of individuals who change their type under the non-monetary incentive – CPR users first visit only

Incentive	Type classification in Baseline	**Type classification from behavior in rounds 11-20**				
		Altruist	Self-regarding	Reciprocator	Counter	Unidentified
Communication one shot	Altruist	44	0	44	0	11
	Self-regarding	0	0	100	0	0
	Reciprocator	26	0	37	0	37
	Counter	50	0	21	14	14
	Unidentified	24	0	29	5	43
Communication each round	Altruist	100	0	0	0	0
	Self-regarding	33	0	33	0	33
	Reciprocator	43	0	38	10	10
	Counter	36	0	29	7	29
	Unidentified	43	0	24	10	24

Table 9b. Percentage of individuals who change their type under the monetary incentive s – CPR users first visit only

Incentive	Type classification in Baseline	Type classification from behavior in rounds 11-20				
		Altruist	Self-regarding	Reciprocator	Counter	Unidentified
Low subsidy	Altruist	100	0	0	0	0
	Self-regarding	44	0	11	22	22
	Reciprocator	60	0	13	7	20
	Counter	50	0	7	14	29
	Unidentified	36	0	29	4	32
Low Fine	Altruist	82	0	18	0	0
	Self-regarding	42	25	0	17	17
	Reciprocator	41	0	30	11	19
	Counter	56	6	13	13	13
	Unidentified	41	0	21	9	29
Medium Fine	Altruist	100	0	0	0	0
	Self-regarding	20	20	60	0	0
	Reciprocator	33	0	33	0	33
	Counter	50	0	0	0	50
	Unidentified	30	10	30	10	20
High Fine	Altruist	100	0	0	0	0
	Self-regarding	50	50	0	0	0
	Reciprocator	50	0	15	10	25
	Counter	50	8	8	25	8
	Unidentified	61	0	13	9	17
Very High Fine	Altruist	50	0	50	0	0
	Self-regarding	0	0	0	0	0
	Reciprocator	67	0	0	0	33
	Counter	43	14	0	29	14
	Unidentified	67	0	0	33	0

In table 9b, a Medium fine, those who were categorized as altruist in the baseline did not change their type with the fine introduction. In Table 9a, we observe that self-regarding

[28] Among other reasons that communication enhances cooperation are: i) it allows players to commit; ii) it allows a process of moralization among players; iii) it can reinforce group identity (Messick and Brewer (1983), Kollock (1998))

individuals disappeared with the use of non-monetary incentives. Only the low fine, medium fine and high fine, 25%, 20%, 50% of those who were classified as self-regarding, kept their title, respectively. Individuals who were classified as counters migrated to other types, mainly altruist and unidentified. In the case of the Low subsidy, those who were categorized as Counter in the baseline, 50% became altruist in treatment phase, 7 % became reciprocator, 29% was unidentified and 14% remained as counter. We observe a change of preferences to altruist. In most treatments, those were classified as Altruist kept their title, although we observe that under communication one shot and the very high fine some who were altruist change their type to reciprocators.

7. Determinants of social preference type

A key feature of heterogeneity is the role of individual socio-economic background and it is relevant to complement the analysis with a comparison between behavior inside and outside the lab, for example, the use of CPR in real life by the participants. Figure 4 shows the fraction of players that extract 8 units according to their economic dependence to the CPR. Those CPR users whose income depends 100% on the CPR extract significantly less whereas those users whose income depends 0% on the CPR extract significantly more[29].

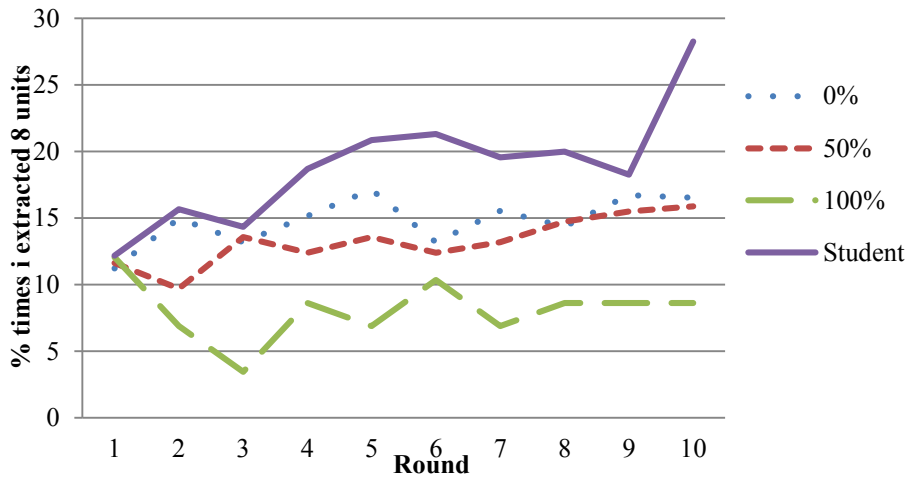


Figure 4. Heterogeneity of real level extraction of the CPR in the game all CPR users vs. students (N=1095). The solid line shows the % time that the Self-regarding NE was chosen in the game by the Students sample. The round-dot line shows the case with individuals who use 0% of the real CPR. The square-dot line shows the average level of extraction in the game by individuals who use 50% of the real CPR. The long-dashed line the average level of extraction in the game by individuals who use 100% of the real CPR. The difference in means in the last round is significant at 10%.

We define $\theta_{zq} = \Theta(Z\phi)$, the vector of conditional preference parameters as a function of some exogenous socio-economic characteristics Z vector[30] and analyze the determinants of being type q in our sample of CPR users in the first visit. In table 10 and 11 we show the role of individual socio-economic characteristics and group composition on the probability of being altruist, self-regarding, reciprocator, and counter, after controlling for a variety of observable variables and using the sample of first visit CPR users only. In Table 10, all eight columns show results obtained from a probit regression with the Index of CPR extraction as

[29] This table shows a similar pattern to Molina (2011) with 665 CPR users.
[30] The vector of unconditional preference parameters is simply θ_q.

main economic activity and in Table 11, all eight columns show results obtained from a probit regression with the Index of time spent by the household on CPR extraction[31].
In order to check whether some variables are near perfect linear combinations of one another, we examine the variance inflation factor (VIF) and the degree of collinearity. The VIF values of these variables are very high, which indicates that these variables are possibly redundant. We exclude these variables on use of the CPR, social capital measures and perceptions from the analysis in Tables 10, 11 and 12. Table E1 in the appendix shows the coefficients obtained in a number of probit regressions that relate the probability that an individual is type q to each of the measures that presents multicollinearity. The purpose of these table is simply to illustrate the association between the behavior in the game we observe and other forms of behavior that have been used in the literature as indicators for pro-social behavior in CPR.

Result 7. *Individual and group level of education increases the likelihood of being an Altruist.*

Result 8. *Perception of lack of willingness to cooperate in the community increases the likelihood of being a reciprocator.*

In tables 10 and 11 in Columns I (1), II (1), III (1) and IV (1) we include socio economic characteristics, such as the age and the education level of the participant (the complete set of results can be found in the Appendix); individual perceptions on the CPR and social capital in the community and an individual measure of social capital, i.e. volunteer work in the previous year. In Columns I (2), II (2), III (2) and IV (2) we add variables on group composition such as the percentage of altruist, reciprocators, self-regarding and counters among the rest of the members in the group, the average sex, years of education and index of economic dependence of the CPR among the rest of the group. Table 11 presents the same analysis but uses an index of time spent by CPR extraction.

Household size, years of education and volunteering increase probability of being categorized as type Altruist; those who in real life depend economically more on the CPR have a lower probability of being allocated to the self-regarding type, confirming our observation from Figure 4. If the participant thinks that the amount of the CPR has remained the same in the past years, the probability of being an altruist decreases 3%, which consistent with anecdotal evidence on people who believe that the resource is not scarce extracting more than the social optimum in order to maximize individual gains. On the other hand, if the participant perceives that there is an interest in the community to cooperate, the probability of this participant being type reciprocator decreases and that of counter increases. Counter individuals would take a higher advantage in a community where people cooperate, as counters will spitefully free-ride in order to obtain a higher benefit. Finally, a participant who thinks the community should be the main vigilant will have a less chance to be self-regarding.

When introducing variables on group composition, we observe that a participant's behavior is conditioned on the types of others in the group (see columns I (2) and I (4). For example, counters were those who extracted a small amount in the first rounds and then extracted a high amount in order to obtain the gain of free riding when the rest of the members were thinking they had mostly altruists in the group. The effect of belonging to a group with a high number of years of education also has a positive effect of behaving as an altruist. In addition,

[31] We report the main results in table 10, table 11 and their additional controls in the appendix.

the average index of CPR extraction as main economic activity among the rest of the group increases the likelihood of behaving as altruist and decreases that of behaving as self-regarding. The result that the percentage of reciprocators in the group lowers the probability of behaving as a self-regarding is consistent: a self-regarding individual who knows she is among reciprocators is more likely to internalize the retaliation costs without further considerations about social preferences. Although being a woman increases the chance to be categorized as self-regarding, the presence of other women in the group reduces it, suggesting a same-sex solidarity that is stronger than self-interest. When we conduct the analysis with the index of time spent on CPR extraction in table 11, the effect of this index with the probability of behaving as self-regarding is no longer significant when controlling for group composition.

Result 9. *Estimated structural parameters confirm results 7 and 8.*

Table 12 presents a linear estimation of the determinants for the individual parameters of altruism and reciprocity, $\hat{\rho}$ and $\hat{\mu}$ respectively, for the first time CPR users. We find a robust positive relation between $\hat{\rho}$ and years of education. Without adding controls on group composition, there is a positive relation between the perception of cooperation in the community and $\hat{\rho}$. We find a robust negative relation between $\hat{\mu}$ and whether the participant perceives there is an interest in the community to cooperate. If the individual perceives her community has an interest for cooperation there is no motivation for including in her utility a component that seeks to enforce a social norm.

Although there is a positive relation between household size, age, the index of time spent on CPR, the index of CPR extraction as main economic activity, these coefficients are no longer significant when including group composition variables. The coefficient for altruism also has a positive relation with the average $\hat{\rho}$, average index of CPR as main economic activity and average index of time spent by CPR extraction among the rest of the group. However, there is no relation between measures of group composition and the coefficient of reciprocity, which may suggest that reciprocity is a behavior that is already embedded before the individual enters the lab, regardless of the composition of the group and how cooperation evolves over time during the game. In contrast, behaving as an altruist or self-regarding is highly related not only to exogenous individual characteristics but also to how other group members behave and how cooperation evolves.

Table 10. Probit estimation of determinants of being type q including the Index of CPR extraction as main economic activity

Dependent variable: 1 if player is type q		*Altruist*		*Self-regarding*		*Reciprocator*		*Counter*	
	Independent Variables	I (1)	I (2)	II (1)	II (2)	III (1)	III (2)	IV (1)	IV (2)
Participant's characteristics	Household Size	0.009*	0.006**	0.002	0.002	-0.004	-0.002	-0.007	-0.009
		[0.005]	[0.003]	[0.003]	[0.003]	[0.010]	[0.010]	[0.009]	[0.009]
	Age	0.000	0.001	-0.001	-0.001*	0.000	0.000	0.001	0.001
		[0.001]	[0.001]	[0.001]	[0.001]	[0.002]	[0.002]	[0.002]	[0.002]
	Sex	0.002	-0.007	0.034*	0.040**	0.052	0.020	-0.059	-0.035
		[0.024]	[0.020]	[0.019]	[0.017]	[0.044]	[0.048]	[0.048]	[0.051]
	Years of education	0.008**	0.004*	-0.004	-0.004	-0.007	-0.003	-0.004	-0.002
		[0.004]	[0.003]	[0.003]	[0.003]	[0.007]	[0.007]	[0.007]	[0.007]
	Index of CPR extraction as main economic activity	0.029	-0.005	-0.099**	-0.067**	-0.097	-0.117	0.077	0.023
		[0.044]	[0.037]	[0.041]	[0.033]	[0.099]	[0.123]	[0.097]	[0.126]
Participant's perceptions	=1 if thinks that the CPR has not changed	-0.043**	-0.029*	-0.008	-0.008	-0.073	-0.085	0.080	0.076
		[0.020]	[0.016]	[0.022]	[0.019]	[0.055]	[0.054]	[0.063]	[0.065]
	Perceived interest in the community to cooperate	0.067	0.039	-0.007	-0.004	-0.193***	-0.190**	0.193**	0.198***
		[0.051]	[0.037]	[0.026]	[0.025]	[0.074]	[0.076]	[0.077]	[0.077]
	=1 if thinks the community should be the main vigilant of the CPR	0.015	0.018	-0.040*	-0.039**	0.028	0.046	0.036	0.030
		[0.026]	[0.020]	[0.021]	[0.019]	[0.046]	[0.046]	[0.045]	[0.046]
Social capital measures	=1 if individual has done any volunteer work in the past year	0.050*	0.040**	-0.021	-0.022	0.035	0.021	-0.049	-0.028
		[0.028]	[0.020]	[0.021]	[0.020]	[0.051]	[0.053]	[0.052]	[0.052]
Group composition (info within group)	% altruists among the rest of the group		0.192***		-0.051		0.149		0.056
			[0.069]		[0.042]		[0.156]		[0.123]
	% reciprocators among the rest of the group		0.111**		-0.074*		-0.083		0.152
			[0.044]		[0.044]		[0.152]		[0.108]
	% counters among the rest of the group		0.126***		-0.028		-0.104		0.304***
			[0.046]		[0.043]		[0.116]		[0.116]
	Average sex among the rest of the group		0.049*		-0.044*		0.071		-0.065
			[0.029]		[0.025]		[0.085]		[0.079]
	Average years of education among the rest of the group		0.015***		-0.004		-0.027**		-0.006
			[0.005]		[0.003]		[0.012]		[0.011]
	Average Index of CPR extraction as main economic activity among the rest of the group		0.126**		-0.064*		-0.151		0.134
			[0.063]		[0.037]		[0.162]		[0.178]
Observations		379	379	379	379	379	379	379	379

* Significant at 10%; ** significant at 5%; *** significant at 1%. Marginal effects reported. Robust standard errors in brackets clustered at the group level. We also control for landownership, if the household's kitchen uses gas or electricity, percentage of self-regarding among the rest of the group and average age among the rest of the group. For the additional controls results, see table E2.

Table 11. Probit estimation of determinants of being type q including the Index of time spent by household on CPR extraction

Dependent variable: 1 if player is type q		*Altruist*		*Self-regarding*		*Reciprocator*		*Counter*	
	Independent Variables	I (3)	I (4)	II (3)	II (4)	III (3)	III (4)	IV (3)	IV (4)
Participant's characteristics	Household Size	0.007*	0.005*	0.003	0.003	-0.005	-0.003	-0.008	-0.009
		[0.004]	[0.003]	[0.003]	[0.003]	[0.010]	[0.010]	[0.010]	[0.010]
	Age	0.001	0.001	-0.001	-0.001*	0.000	0.000	0.001	0.001
		[0.001]	[0.001]	[0.001]	[0.000]	[0.002]	[0.002]	[0.002]	[0.002]
	Sex	0.003	-0.005	0.032*	0.037**	0.050	0.007	-0.056	-0.020
		[0.024]	[0.021]	[0.018]	[0.016]	[0.044]	[0.049]	[0.049]	[0.049]
	Years of education	0.008**	0.005*	-0.004	-0.004*	-0.006	-0.002	-0.005	-0.003
		[0.004]	[0.003]	[0.003]	[0.002]	[0.007]	[0.007]	[0.007]	[0.007]
	Index of time spent by the household on CPR extraction	0.082	0.049	-0.141***	-0.074	-0.057	-0.234	0.110	0.239
		[0.065]	[0.044]	[0.051]	[0.062]	[0.123]	[0.158]	[0.113]	[0.153]
Participant's perceptions	=1 if thinks that the CPR has not changed	-0.047**	-0.031*	-0.007	-0.007	-0.076	-0.096*	0.077	0.084
		[0.019]	[0.017]	[0.021]	[0.018]	[0.054]	[0.053]	[0.063]	[0.065]
	Perceived interest in the community to cooperate	0.061	0.044	-0.009	-0.005	-0.190***	-0.193***	0.193**	0.203***
		[0.051]	[0.039]	[0.024]	[0.022]	[0.073]	[0.073]	[0.078]	[0.073]
	=1 if thinks the community should be the main vigilant of the CPR	0.016	0.020	-0.038*	-0.037**	0.026	0.044	0.037	0.024
		[0.026]	[0.021]	[0.020]	[0.018]	[0.047]	[0.047]	[0.045]	[0.047]
Social capital measures	=1 if individual has done any volunteer work in the past year	0.052*	0.041**	-0.015	-0.019	0.038	0.030	-0.049	-0.039
		[0.027]	[0.020]	[0.018]	[0.019]	[0.050]	[0.052]	[0.051]	[0.053]
Group composition (info within group)	% altruists among the rest of the group		0.173***		-0.024		0.131		0.070
			[0.063]		[0.040]		[0.143]		[0.122]
	% reciprocators among the rest of the group		0.118**		-0.067*		-0.048		0.132
			[0.047]		[0.040]		[0.149]		[0.110]
	% counters among the rest of the group		0.141***		-0.040		-0.112		0.314***
			[0.052]		[0.040]		[0.115]		[0.115]
	Average sex among the rest of the group		0.039		-0.042*		0.081		-0.079
			[0.030]		[0.024]		[0.086]		[0.077]
	Average years of education among the rest of the group		0.015***		-0.005		-0.023*		-0.010
			[0.005]		[0.003]		[0.012]		[0.011]
Observations		379	379	379	379	379	379	379	379

* Significant at 10%; ** significant at 5%; *** significant at 1%. Marginal effects reported. Robust standard errors in brackets clustered at the group level. We also control for landownership, if the household's kitchen uses gas or electricity, percentage of self-regarding among the rest of the group and average age among the rest of the group. For the additional controls results, see table E3.

Table 12. Linear estimation of determinants of parameters $\hat{\rho}$ and $\hat{\mu}$. First visit CPR users.

	Dependent variable	$\hat{\rho}$				$\hat{\mu}$			
	Independent Variables	V (1)	V (2)	V (3)	V (4)	VI (1)	VI (2)	VI (3)	VI (4)
Participant's characteristics	Household Size	0.019***	0.005	0.018***	0.003	0.002	0.002	0.002	0.001
		[0.006]	[0.005]	[0.006]	[0.005]	[0.002]	[0.002]	[0.002]	[0.002]
	Age	0.006***	0.001	0.006***	0.001	0.000	-0.000	0.000	-0.000
		[0.001]	[0.001]	[0.001]	[0.001]	[0.000]	[0.000]	[0.000]	[0.000]
	Years of education	0.029***	0.011**	0.028***	0.012**	0.000	0.000	0.000	0.000
		[0.005]	[0.005]	[0.005]	[0.005]	[0.001]	[0.002]	[0.001]	[0.001]
	Index of CPR extraction as main economic activity	0.243***	-0.013			-0.010	-0.012		
		[0.064]	[0.068]			[0.021]	[0.030]		
	Index of time spent by the household on CPR extraction			0.273***	-0.067			-0.003	-0.051
				[0.085]	[0.098]			[0.027]	[0.038]
Participant's perceptions	=1 if thinks that the CPR has not changed	-0.030	-0.044	-0.031	-0.046	-0.021	-0.021	-0.022	-0.025*
		[0.052]	[0.047]	[0.053]	[0.047]	[0.015]	[0.014]	[0.015]	[0.014]
	Perceived interest in the community to cooperate	0.184***	0.040	0.178***	0.043	-0.046***	-0.051***	-0.046***	-0.053***
		[0.061]	[0.056]	[0.062]	[0.057]	[0.016]	[0.017]	[0.016]	[0.016]
Social capital measures	=1 if individual has done any volunteer work in the past year	0.044	-0.011	0.042	-0.009	0.019	0.011	0.019*	0.014
		[0.033]	[0.031]	[0.035]	[0.031]	[0.012]	[0.012]	[0.011]	[0.012]
Group composition (info within group)	Average $\hat{\rho}$ among the rest of the group		0.469***		0.466***		0.010		-0.001
			[0.079]		[0.066]		[0.023]		[0.023]
	Average age among the rest of the group		0.003		0.002		0.001*		0.001*
			[0.002]		[0.002]		[0.001]		[0.001]
	Average years of education among the rest of the group		0.020***		0.020***		-0.002		-0.001
			[0.007]		[0.007]		[0.002]		[0.002]
	Average Index of CPR extraction as main economic activity among the rest of the group		0.216**				-0.032		
			[0.087]				[0.041]		
	Average Index of time spent on CPR extraction among the rest of the group				0.329***				0.048
					[0.106]				[0.042]
Observations		379	379	379	379	379	379	379	379
R-squared		0.873	0.900	0.872	0.901	0.058	0.088	0.058	0.089

* Significant at 10%; ** significant at 5%; *** significant at 1%. Robust standard errors in brackets clustered at the group level. We also control for sex, landownership, if the household's kitchen uses gas or electricity, average sex among the rest of the group and average $\hat{\mu}$ among the rest of the group. For the additional controls results, see table E4.

8. Concluding remarks

What is the role of the existence of different social preferences within a community in overcoming social dilemmas or the social efficiency of economic incentives? This question is not easily answered because differences in observed behavior may have two sources. People may differ in socio-economic characteristics as well as in perceptions and beliefs on the Common Pool Resource (CPR) and social capital within the community, which are exogenous for the policy designer. Yet another source of difference is the composition of social preferences in the surrounding community, which is revealed in the day-by-day interaction related to the CPR extraction process.

This is a study on type classification of social preferences (i.e. unconditional altruist, self-regarding, reciprocator and inequity averse) within a CPR game we conducted in some Colombian rural areas. We use a RCM specification to identify individual types and at the same time estimate the theoretical parameters within a (theory-based) structural specification. Within the same specification, we attempt to understand potentially different utility-maximizing behaviors across types. We find that individual pro-social behavior is consistently different across types: there are altruists who extract less, self-regarding individuals who extract more, and reciprocators who behave according to what we interpret to be a group norm. The composition of types in the group significantly explains the average level of extraction prior to treatment, the path of extraction and the outcomes.

In addition, we examine the impact of the incentive on each type and the composition of the group on the social efficiency of incentives. We find that different incentives have a different effect on behavior and non-monetary incentives are more effective in groups where other-regarding preferences are prevalent, whereas only the subsidy is effective in promoting behavior among self-regarding individuals and negative reciprocators (i.e. those who respond to individuals extracting an amount higher than the social norm). Estimating the RCM classification again after the incentives are introduced, we find that self-regarding behavior (own-profit-maximizing) endures partially over time: many self-regarding individuals move away from full extraction to minimal extraction (the new NE under high/very high fine) as the fine increases, consistent with own profit maximizing behavior. However, suboptimally high levels of extraction remain within this type, suggesting that among the Self-regarding category there are irrational over-extractors. Our specification generates a corner solution for the self-regarding NE. In other to identify over-extractors from rational maximizers, interior solutions for the best reply functions are needed.[32] The very high fine affects altruists since half of them become reciprocators. In sum, incentives may alter social preferences in forms that claim further exploration, confirming the results surveyed in the first chapter of this dissertation (Bowles and Polania-Reyes (2012)).

In order to compare our experimental results with real life behavior, we perform a probit regression analysis of individual type classification and estimate a linear model for both the altruism and the reciprocity parameters. Both analyses set these key variables as a function of socioeconomic characteristics, perceptions, social capital within the community and group composition (in terms of types and the observable variables). The results suggest that behavior may be determined, in part, by the composition of types (and their demographic characteristics) in the group as well as the perceived scarcity of the CPR, the level of

[32] We conjecture $U_i^S = \pi_i^{1/2}$ might work.

community willingness to cooperate perceived, whether the participant does volunteer work and whether the CPR is the household's main economic activity.

Other extensions are relevant for the discussion of heterogeneity. First and most important, this study motivates research that aims to identify a distribution of types -which are not observable- and not just point estimates as inferred from a random coefficients model. This is possible by using a finite mixtures model (for an example, see Echeverry and Polania-Reyes (2015), Bruhin, et al. (2010), Coller et al. (2011), Conte, et al. (2011) and Cappelen et al. (2013)).[33]

We examine the most popular types of social preferences in the literature. We also provide indicative evidence of determinants of altruistic or reciprocal behavior. Our model is nonlinear in parameters due to the presence of the β parameter of inequity aversion. The RCM approach does not allow for this non-linearity to identify the set of four types simultaneously given that we do not have enough exclusion restrictions. Although our cross-sectional specification relies on four specific types, testing for the optimal number of types would help determine whether the profiles chosen are appropriate and verify that the correct model specification is being used.

A trade-off seems to stem from two possible approaches. If types are derived from observing a short timeframe (even from one-shot games, e.g. Cappelen et al. (2013)) the evolution of the type can then be tracked over time but the question arises as to the robustness of identification. If instead types are defined from observing a consistent behavior over time, then the definition is robust but the apparition and evolution of types is out of question. Identification, which is the key issue in the study of social preferences, has therefore a time component to it as well as the (previously discussed) cross-sectional component. Addressing identification over the time series component likely calls for more sophisticated econometric techniques. Coupled with further studies on the underlying social and psychological mechanisms, such research can explain the formation of social preferences within a structural model.

9. References

Ahn, T.K., Ostrom, E., and Walker, J. 2011. A Common-Pool Resource Experiment with Postgraduate Subjects from 41 Countries. Ecological Economics, 70 (9). Pp. 1580-1589.

Akerlof, G. A. 1984. An Economic Theorist's Book of Tales. Cambridge, UK: Cambridge University Press.

Andersen, S., G. Harrison, M. Lau, and E. Rutström. 2010. Preference heterogeneity in experiments: Comparing the field and laboratory. Journal of Economic Behavior and Organization 73 (2), 209-224.

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

Arrow, K. J. 1971. Political and Economic Evaluation of Social Effects and Externalities, pp. 3-23. Amsterdam: North Holland.

Bardsley, N. and P. Moffatt. 2007. The experimetrics of public goods: Inferring motivations from contributions. Theory and Decision 62 (2), 161-193.

[33] This data set presents a challenge in term of estimation of finite mixtures since we need to specify a likelihood function that takes into account the panel structure of the data.

Becker, G. 1976. Altruism, egoism, and genetic fitness: Economics and sociobiology. Journal of Economic Literature. 14 (3), 817-826.
Bewley, T. 1999. Why wages don't fall during a recession. Harvard Univ Press.
Bicchieri, C. 2006 The Grammar of Society: the Nature and Dynamics of Social Norms, New York: Cambridge University Press.
Bicchieri C. 2014. Norms in the Wild: How to Diagnose, Measure and Change Social Norms. Cambridge: Cambridge University Press.
Bicchieri, C. and Muldoon, Ryan, 2014 "Social Norms", The Stanford Encyclopedia of Philosophy (Spring Edition), E. N. Zalta (ed.), URL: http://plato.stanford.edu/archives/spr2014/entries/social-norms
Bohnet, I. and S. Huck. 2004. Repetition and reputation: Implications for trust and trustworthiness when institutions change. The American Economic Review. 94 (2), 362-366.
Bolton, G. E. and A. Ockenfels. 2000. Erc: A theory of equity, reciprocity, and competition. The American Economic Review. 90 (1), 166-193.
Bowles, S. 2004. Microeconomics: Behavior, Institutions, and Evolution. Princeton: Princeton University Press.
Bowles, S. 2008. Policies designed for self-interested citizens may undermine "the moral sentiments": Evidence from economic experiments. Science 320 (5883), 1605-1609.
Bowles, S. and H. Gintis. 2003. Strong reciprocity: The origin of human cooperation. Cambrige: MIT Press.
Bowles, S. and S. Polania-Reyes. 2012. Economic incentives and social preferences: Substitutes or complements? Journal of Economic Literature. 50 (2), 368–425.
Bruhin, A., H. Fehr-Duda, and T. Epper. 2010. Risk and rationality: Uncovering heterogeneity in probability distortion. Econometrica 78 (4), 1375-1412.
Brosig, Jeannete. 2002 'Identifying Cooperative Behavior: Some Experimental Results in a Prisoner's Dilemma Game', Journal of Economic Behavior and Organization, 47, 275–90.
Burlando, R. and F. Guala, 2005. Heterogeneous agents in public goods experiments. Experimental Economics 8 (1), 35-54.
Cappelen, A. W., Moene, K. O., Sørensen, E. Ø. and Tungodden, B. 2013. Needs versus entitlements—an international fairness experiment. Journal of the European Economic Association, 11: 574–598. doi: 10.1111/jeea.12000
Cárdenas, J.C, Stranlund, J. and Willis, C. 2000. "Local Environmental Control and Institutional Crowding-Out," World Development, Elsevier, vol. 28(10), p. 1719-1733, October.
Cárdenas, J. C., & Carpenter, J. P. 2005. Three themes on field experiments and economic development. Research in Experimental Economics, 10, 71-123.
Cárdenas, J. C. 2011. Social norms and behavior in the local commons as seen through the lens of field experiments. Environmental and Resource Economics 48 (3), 451-485.
Cárdenas, J.C., C. Mantilla and R. Sethi. 2015. Stable Sampling Equilibrium in Common Pool Resource Games (June). At: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2304566
Cárdenas, J. C. 2004. Norms from outside and inside: an experimental analysis on the governance of local ecosystems. Forest Policy and Economics 6, 229-241.
Cárdenas, JC., T. K. Ahn, and E. Ostrom. 2004 "Communication and cooperation in a common pool resource dilemma: a field experiment." Advances in Understanding Strategic Behaviour: Game Theory, Experiments and Bounded Rationality: 258-86.
Cardenas, J C. 2005 "Groups, commons and regulations: experiments with villagers and students in Colombia." In Psychology, Rationality and Economic Behaviour: Challenging

Standard Assumptions, Bina Agarwal and Alessandro Vercelli (eds.). Palgrave Macmillan (in association with the International Economic Association), New York, IEA Conference Volume Series. Vol. 142.
Carpenter, J., S. Bowles, H. Gintis, and S.-H. Hwang. 2009. Strong reciprocity and team production: Theory and evidence. Journal of Economic Behavior and Organization 71 (2), 221-232.
Carpenter, J. P. and E. Seki. 2011. Do social preferences increase productivity? Field experimental evidence from fishermen in Toyama bay. Economic Inquiry. 49(2), 612–630.
Casari, M. and C. Plott. 2003. Decentralized management of common property resources: experiments with a centuries-old institution. Journal of Economic Behavior and Organization 51 (2), 217-247.
Charness, G. and M. Rabin. 2002. Understanding social preferences with simple tests*. Quarterly Journal of Economics 117 (3), 817-869.
Coller, M., G. Harrison, and E. Rutström. 2011. Latent process heterogeneity in discounting behavior. Oxford Economic Papers.
Conte, A., J. D. Hey, and P. G. Moffatt. 2011. Mixture models of choice under risk. Journal of Econometrics 162 (1), 79-88.
Dufwenberg, M., & Kirchsteiger, G. 2004. A theory of sequential reciprocity. Games and economic behavior, 47(2), 268-298.
Dufwenberg, M., P. Heidhues, G. Kirchsteiger, F. Riedel, and J. Sobel. 2011. Other-regarding preferences in general equilibrium. Review of Economic Studies 78 (2), 613.
Duffy, John, and Félix Muñoz-García. 2013 "Cooperation and signaling with uncertain social preferences." Theory and Decision 78.1: 45-75.
Elster, J. 1998. Emotions and economic theory. Journal of Economic Literature 36 (1), 47-74.
Falk, A., E. Fehr, and U. Fischbacher. 2002. Appropriating the Commons: A Theoretical Explanation. In: Ostrom, Elinor, et al., eds. The drama of the commons. National Academies Press, 157-191.
Falk, A., E. Fehr, and U. Fischbacher. 2005. Driving forces behind informal sanctions. Econometrica 73 (6), 2017-2070.
Falk, A., & Fischbacher, U. 2006. A theory of reciprocity. Games and Economic Behavior, 54(2), 293-315.
Falk, A. and J. J. Heckman. 2009. Lab experiments are a major source of knowledge in the social sciences. Science 326 (5952), 535-538.
Falk, A. and M. Kosfeld. 2006. The hidden costs of control. American Economic Review 96 (5), 1611-1630.
Fehr, E. and S. Gächter. 2000. Cooperation and punishment in public goods games. American Economic Review 90 (4), 980-994.
Fehr, E., S. Gächter, and G. Kirchsteiger. 1997. Reciprocity as a contract enforcement device: Experimental evidence. Econometrica 65 (4), 833-860.
Fehr, E. and K. M. Schmidt. 1999. A theory of fairness, competition, and cooperation. Quarterly Journal of Economics 114 (3), 817-868.
Fischbacher, U., S. Gächter, and E. Fehr. 2001. Are people conditionally cooperative? Evidence from a public goods experiment. Economics Letters 71 (3), 397-404.
Frank, Robert H. 1988. Passions within Reason: The Strategic Role of the Emotions, New York, W. W. Norton.
Hackett, S., Schlager, E. and Walker, J. M. 1994. The Role of Communication in Resolving Commons Dilemmas: Experimental Evidence with Heterogeneous Appropriators. Journal of Environmental Economics and Management, 27 (2), 99-126.
Holms, T. H., E. Kjerstad, H. Lurs, and O. R. Straume. 2010. Does monetary punishment

crowd out pro-social motivation? a natural experiment on hospital length of stay. Journal of Economic Behavior and Organization 75 (2), 261-267.
Hwang, S.-H. and S. Bowles. 2014. Optimal incentives with state-dependent preferences. Journal of Public Economic Theory, 16 (5), 681-705
Janssen, M. A., R. Holahan, A. Lee, and E. Ostrom. 2010. Lab experiments for the study of social-ecological systems. Science 328 (5978), 613-617.
Kessler, E. 2008. Behavioural Economics of Performance Incentives. Ph. D. thesis.
Kollock, Peter, 1998 'Social Dilemmas: The Anatomy of Cooperation', Annual Review of Sociology, 24, 183–214.
Kreps, D. M. 1997. Intrinsic motivation and extrinsic incentives. The American Economic Review 87 (2), 359-364.
Kreps, D. M., P. Milgrom, J. Roberts, and R. Wilson. 1982. Rational cooperation in the finitely repeated prisoner's dilemma. Journal of Economic Theory 27 (2), 245-252.
Kurzban, R. and D. Houser. 2001. Individual differences in cooperation in a circular public goods game. European Journal of Personality 15 (S1), S37-S52.
Kurzban, R. and D. Houser. 2005. Experiments investigating cooperative types in humans: A complement to evolutionary theory and simulations. Proceedings of the National Academy of Sciences of the United States of America 102 (5), 1803-1807.
Ledyard, J. 1995. Public goods: a survey of experimental research, in ''Handbook of Experimental Economics' J. Kagel and A. Roth, Eds.., Princeton Univ. Press, Princeton, NJ.
Leider, S., M. M. Mbius, T. Rosenblat, and Q.-A. Do. 2009. Directed altruism and enforced reciprocity in social networks. The Quarterly Journal of Economics 124 (4), 1815-1851.
Levine, D. K. 1998. Modeling altruism and spitefulness in experiments. Review of Economic Dynamics 1 (3), 593-622. Department of Economics, UCLA.
Loewenstein, G. F., L. Thompson, and M. H. Bazerman. 1989. Social utility and decision making in interpersonal contexts. Journal of Personality and Social Psychology 57 (3), 426-441.
Mailath, G. and L. Samuelson. 2006. Repeated games and reputations: long-run relationships. Oxford University Press, USA.
Manski, C. 2011. Policy analysis with incredible certitude*. The Economic Journal 121 (554), F261-F289.
Margreiter, M., M. Sutter, and D. Dittrich. 2005. Individual and collective choice and voting in common pool resource problem with heterogeneous actors. Environmental and Resource Economics 32 (2), 241-271.
Messick, D. M. and Brewer, M. B. (1983) 'Solving Social Dilemmas: A Review', Review of Personality and Social Psychology, 4, 11–33.
Molina, A. 2011. ¿Sobreexplotación o conservación?: evidencia experimental sobre un problema de dependencia económica en comunidades rurales. Coyuntura Económica, [Over explotation or conservation? Experimental evidence about an issue of finantial dependency in rural communities] 12,(1) June, pp. 139-183.
Ostrom, E. 1990. Governing the Commons: The Evolution of Institutions for Collective Action. Cambridge, UK: Cambridge University Press.
Ostrom, E., Roy Gardner and James Walker, 1994 Rules, games and common pool resources. Ann Arbor: University of Michigan Press.
Polania-Reyes, S and D. Echeverry, 2015. Identification of Other-regarding Preferences:
Evidence from a Common Pool Resource game in Colombia. Mimeo
Rabin, M. 1993. Incorporating fairness into game theory and economics. American Economic Review 83 (5), 1281-1302.
Rabin, M. 1998. Psychology and economics. Journal of Economic Literature 36 (1), 11-46.

Rodriguez-Sickert, C., R. A. Guzmán, and J. C. Cárdenas. 2008. Institutions influence preferences: Evidence from a common pool resource experiment. Journal of Economic Behavior and Organization 67 (1), 215-227.
Rustagi, D., S. Engel, and M. Kosfeld. 2010. Conditional cooperation and costly monitoring explain success in forest commons management. Science 330 (6006), 961-965.
Sobel, J. 2005. Interdependent preferences and reciprocity. Journal of Economic Literature 43 (2), 392-436.
Swamy, P. 1970. Efficient inference in a random coefficient regression model. Econometrica: Journal of the Econometric Society, 311-323.
Velez, M. A., J. K. Stranlund, and J. J. Murphy. 2009. What motivates common pool resource users? Experimental evidence from the field. Journal of Economic Behavior and Organization 70 (3), 485-497.
Walker, J. M., R. Gardner, and E. Ostrom. 1990. Rent dissipation in a limited-access common-pool resource: Experimental evidence. Journal of Environmental Economics and Management 19 (3), 203-211.
Zhosan, D. and R. Gardner, 2013. Problems of the commons: group behavior, cooperation and sanctioning in a two-harbor experiment. Environment and Development Economics, 18, pp 7-25

Appendix A. Experimental protocol

Table A1. Table points of the Common Pool Resource game.

Total level of extraction by others	My level of extraction from the resource								Average level of extraction by others
	1	2	3	4	5	6	7	8	
4	758	790	818	840	858	870	878	880	**1**
5	738	770	798	820	838	850	858	860	**1**
6	718	750	778	800	818	830	838	840	**2**
7	698	730	758	780	798	810	818	820	**2**
8	678	710	738	760	778	790	798	800	**2**
9	658	690	718	740	758	770	778	780	**2**
10	638	670	698	720	738	750	758	760	**3**
11	618	650	678	700	718	730	738	740	**3**
12	598	630	658	680	698	710	718	720	**3**
13	578	610	638	660	678	690	698	700	**3**
14	558	590	618	640	658	670	678	680	**4**
15	538	570	598	620	638	650	658	660	**4**
16	518	550	578	600	618	630	638	640	**4**
17	498	530	558	580	598	610	618	620	**4**
18	478	510	538	560	578	590	598	600	**5**
19	458	490	518	540	558	570	578	580	**5**
20	438	470	498	520	538	550	558	560	**5**
21	418	450	478	500	518	530	538	540	**5**
22	398	430	458	480	498	510	518	520	**6**
23	378	410	438	460	478	490	498	500	**6**
24	358	390	418	440	458	470	478	480	**6**
25	338	370	398	420	438	450	458	460	**6**
26	318	350	378	400	418	430	438	440	**7**
27	298	330	358	380	398	410	418	420	**7**
28	278	310	338	360	378	390	398	400	**7**
29	258	290	318	340	358	370	378	380	**7**
30	238	270	298	320	338	350	358	360	**8**
31	218	250	278	300	318	330	338	340	**8**
32	198	230	258	280	298	310	318	320	**8**

Period	-2 to 0	1 to 10	11 to 20	
	Practice	Baseline Open access	Treatment Incentive	Payment + Survey

Figure Error! No text of specified style in document.**1 Timeline of the CPR game**

i. Experimental stages

The following stages were conducted for each of the sessions. This is cited by Cárdenas (2011).

Pre-game Stage (Instructions and Practice Rounds)
Each session of an experiment began with the welcoming and reading of the instructions to the group of five players, as well as the handing out of the following forms (available from author): the GAME CARDS, where participants wrote their choice for every round, that is, their extraction level; the DECISIONS RECORDS SHEET, where participants kept records of their choices and earnings; and the PAYOFF TABLE (see Table A1). Once all questions from participants were clarified, the experimenter continued by conducting one or two practice rounds as examples (see Figure **Error! No text of specified style in document.**1). After resolving all outstanding questions, stage 1 began.

Stage 1 (Rounds 1 to 10)
In Stage 1 of the experiment, each of the players had to decide privately their individual level of extraction from the commons. The decision was written down on one yellow slip (game card); the same information was also recorded on the blue records sheet. The monitor collected the five slips, added the total extraction for the group, which he wrote on the monitor's record sheet, and then announced publicly the total. Each player had to write down the group's total; by subtracting his or her individual extraction, the player was able to calculate his or her payoff for that round using the payoffs table. The player then wrote his or her total gains for the round and the experiment proceeded to the next round with the filling of a new slip. Under such rounds, it was common information that round 10 was the final round. Once they had finished calculating their earnings for round 10, players were told that the rules of the exercise were going to change for Stage 2 of the game. Additionally, they were never told in advance what the rules for Stage 2 were.

Stage 2 (New Rules, Rounds 11 to 20)
The second stage began with the announcement that they would be playing another 10 rounds under a new set of rules. For this stage, the previous record sheets were collected and new ones were distributed among the five players. For the case of face-to-face communication, we began Stage 2 by indicating to the participants that in every round, and prior to their making their decisions, they would be allowed to have a 3 to 5 min discussion on anything they wanted concerning the developments of the game, though no arrangements would be allowed for redistributing earnings once the experiment had ended. However, they were told that decisions would remain private and confidential. For the groups under the regulation treatments, Stage 2 began with an explanation from the experimenter in the following terms. The experimenter reminded the group that they had probably noticed that the group could earn the maximum of points if every player chose a level of extraction equal to one unit (this information was not given to the communication groups however).
They were also told that for achieving such a goal, the monitor would choose one player randomly for every round, and would verify his or her compliance with the stated rule. The

probability of such inspection was of 0.2, and was conducted by drawing a ball with a number from 5 balls in a bag. If a player was inspected and had chosen a higher level of extraction, his or her earnings were reduced by $50 ($175 for the high penalty treatment) times the units of extraction above 1. In the case where there was no fine, the monitor announced publicly the extraction level of the randomly chosen player, and proceeded to the next round. We also had control groups under a baseline treatment, with no change in the rules for Stage 2[34].

The text of the rule is the following: "You may have noticed that if each player in the group chooses a level of extraction of 1 unit the group makes the maximum possible of points. With this rule we will try that the group earns the maximum possible. We will try with this rule that each player in your group chooses a LEVEL OF EXTRACTION of 1 unit."

The Exit Stage (Calculating Earnings, Filling Out the Survey)
Following all of the rounds from Stage 2, the monitors calculated the total earnings for each player by adding the column of round earnings and subtracting the cases where a fine was imposed. While the monitors made the calculations, the players responded to the exit survey, anonymously and in private. Upon returning the filled survey, payments were made in cash to each player and in private.

ii. Experiment instructions (English translation) (from Cardenas (2005:268-265)

These instructions were originally written in Spanish and translated from the final version used in the field work. The instructions were read to the participants from the script below by the same person during all sessions. The participants could interrupt and ask questions at any time. Whenever the following type of text and font e.g. [. . . MONITOR: distribute PAYOFFS TABLE to participants . . .] is found below, it refers to specific instructions to the monitor at that specific point; when in italics, these are notes added to clarify issues to the reader.
Neither of these were read to participants. Where the word 'poster' appears, it refers to a set of posters we printed in very large format with the payoffs table, forms, and the three examples described in the instructions. These posters were hung on a wall near the participants' desks where the eight people could see them easily.

COMMUNITY RESOURCES GAME (Instructions)

Greetings. . .
We want to thank everyone here for attending the call, and specially thank the field practitioner _____ (name of the contact person in that community), and _____ (local organization that helped in the logistics) who made this possible. We will spend about two hours between explaining the exercise, playing it and finishing with a short survey at the exit. So, let us get started.
The following exercise is a different and entertaining way of participating actively in a project about the economic decisions of individuals. Besides participating in the exercise, and being able to earn some prizes and some cash, you will participate in a community workshop

[34] The reason for announcing this was to make sure that the players had a benchmark with which to compare when facing a penalty if chosen for inspection; also to ensure that the external policy was common knowledge. For many sessions, it was very clear that, by round 10 of the first stage, this was the social optimum solution for many of the players. On no single occasion was such a solution questioned, although participants were not allowed to formulate questions prior to stage 2.

in two days to discuss the exercise and other matters about natural resources. During the day of the workshop we will give you what you earn during the game. The funds to cover these expenditures have been donated by various international organizations and the University.

Introduction

This exercise attempts to recreate a situation where a group of families must make decisions about how to use the resources of, for instance, a forest, a water source, a mangrove, a fishery, or any other case where communities use a natural resource. In the case of this community _____ (name of the specific village), an example would be the use of firewood or logging in the _____ (name of an actual local commons area in that village) zone. You have been selected to participate in a group of five people among those that signed up for playing. The game in which you will participate now is different from the ones others have already played in this community, thus, the comments that you may have heard from others do not apply necessarily to this game. You will play for several rounds equivalent, for instance, to years or harvest seasons. At the end of the game you will be able to earn some prizes in kind and cash. The cash prizes will depend on the quantity of points that you accumulate after several rounds.

The PAYOFFS TABLE

To be able to play you will receive a PAYOFFS TABLE equal to the one shown in the poster. [. . . MONITOR: show PAYOFFS TABLE in poster and distribute PAYOFFS TABLE to participants. . .]

This table contains all the information that you need to make your decision in each round of the game. The numbers that are inside the table correspond to points (or pesos) that you would earn in each round. The only thing that each of you has to decide in each round is the LEVEL OF EXTRACTION that you want to allocate extracting resources (in the columns from 1 to 8).

To play in each round you must write your decision number between 1 and 8 in a yellow GAME CARD like the one I am about to show you. [. . . MONITOR: show yellow GAME CARDS and show in the poster. . .] It is very important that we keep in mind that the decisions are absolutely individual, that is, that the numbers we write in the game card are private and that we do not have to show them to the rest of members of the group if we do not want to. The monitor will collect the 5 cards from all participants, and will add the total units of extraction that the group decided to allocate. When the monitor announces the group total, each of you will be able to calculate the points that you earned in the round. Let us explain this with an example. In this game we assume that each player extract as maximum of 8 units of a resource like firewood or logs. In reality this number could be larger or smaller but for purposes of our game we will assume 8 as maximum. In the PAYOFFS TABLE this corresponds to the columns from 1 to 8. Each of you must decide from 1 to 8 in each round. But to be able to know how many points you earned, you need to know the decisions that the rest in the group made. That is why the monitor will announce the total for the group in each round. For instance, if you decide to extract 2 units and the rest of the group together, add to 20 units, you would gain ____ points. Let us look at two other examples in the poster.

[. . .MONITOR: show poster with the THREE EXAMPLES…]

Let us look how the game works in each round.

The DECISIONS FORM

To play each participant will receive one green DECISIONS FORM like the one shown in the poster in the wall. We will explain how to use this sheet. [. . .MONITOR: show the DECISIONS FORM in the poster and distribute the DECISIONS FORMS. . .]

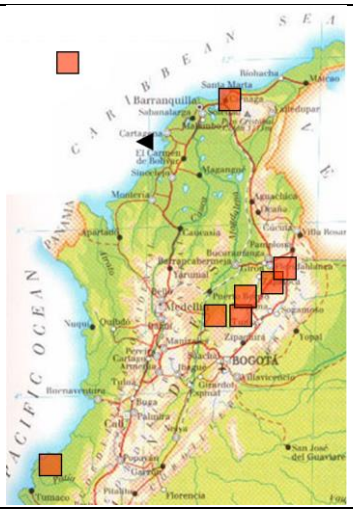
With the same examples, let us see how to use this DECISIONS FORM. Suppose that you decided to play 5 units in this round. In the yellow GAME CARD you should write 5. Also you must write this number in the first column A of the decisions form. The monitor will collect the 5 yellow cards and will add the total of the group. Suppose that the total added 26 units. Thus, we write 26 in the column B of the decisions form. [. . .MONITOR: In the poster, write the same example numbers in the respective cells. . .] To calculate the third column (C), we subtract from the group total, MY DECISION and then we obtain THEIR LEVEL OF EXTRACTION which we write in column C. In our example, 26 − 5 = 21. If we look at the PAYOFFS TABLE, when MY EXTRACTION are 5 and THEIR EXTRACTION are 21, I earn ____ points. I write then this number in the column D of the DECISIONS FORM.
It is very important to clarify that nobody, except for the monitor, will be able to know the number that each of you decides in each round. The only thing announced in public is the group total, without knowing how each participant in your group played. Let us repeat the steps with a new example. [. . .MONITOR: Repeat with the other two examples, writing the numbers in the posters hanging in the wall. . .] It is important to repeat that your game decisions and earnings information are private. Nobody in your group or outside of it will be able to know how many points you earned or your decisions during rounds. We hope these examples help you understand how the game works, and how to make your decisions to allocate your UNITS OF EXTRACTION in each round of the game. If at this moment you have any question about how to earn points in the game, please raise your hand and let us know.
[. . .MONITOR: pause to resolve questions. . .]
It is very important that while we explain the rules of the game you do not engage in conversations with other people in your group. If there are no further questions about the game, then we will assign the numbers for the players and the rest of forms needed to play.

Preparing for playing
Now write down your player number in the green DECISIONS FORM. Write also the place _____ and the current date and time __/__/__, __:__am/pm. In the following poster we summarize for you the steps to follow to play in each round. Please raise your hand if you have a question. [MONITOR: Read the steps to them from the poster] Before we start, and once all players have understood the game completely, the monitor will announce one additional rule for this group. To start the first round of the game we will organize the seats and desks in a circle where each of you face outwards. The monitor will collect your yellow game cards in each round. Finally, to get ready to play the game, please let us know if you have difficulties reading or writing numbers. If so, one of the monitors will sit next to you and assist you with these. Also, please keep in mind that from now on there should be no conversation nor should statements be made by you during the game, unless you are allowed to. We will first have a few rounds of practice that will NOT count toward your real earnings, they are just for practicing the game.

Appendix B. Labs in the field Data

Table B1. Labs in the field: geographical allocation

Villages	CPR
Providencia	Coral reefs Coastal fisheries Crab gatherers
Gaira	Coastal fisheries
Sanquianga	Clamps Fisheries Shrimp Mangroves
Barichara	Andean Forests
Chaina	Firewood
Tabio	Andean Forests Water
La Vega	Water
Neusa	Damn reservoir Trout fishing



Note: The red squares are the villages. Source: Cárdenas (2011)

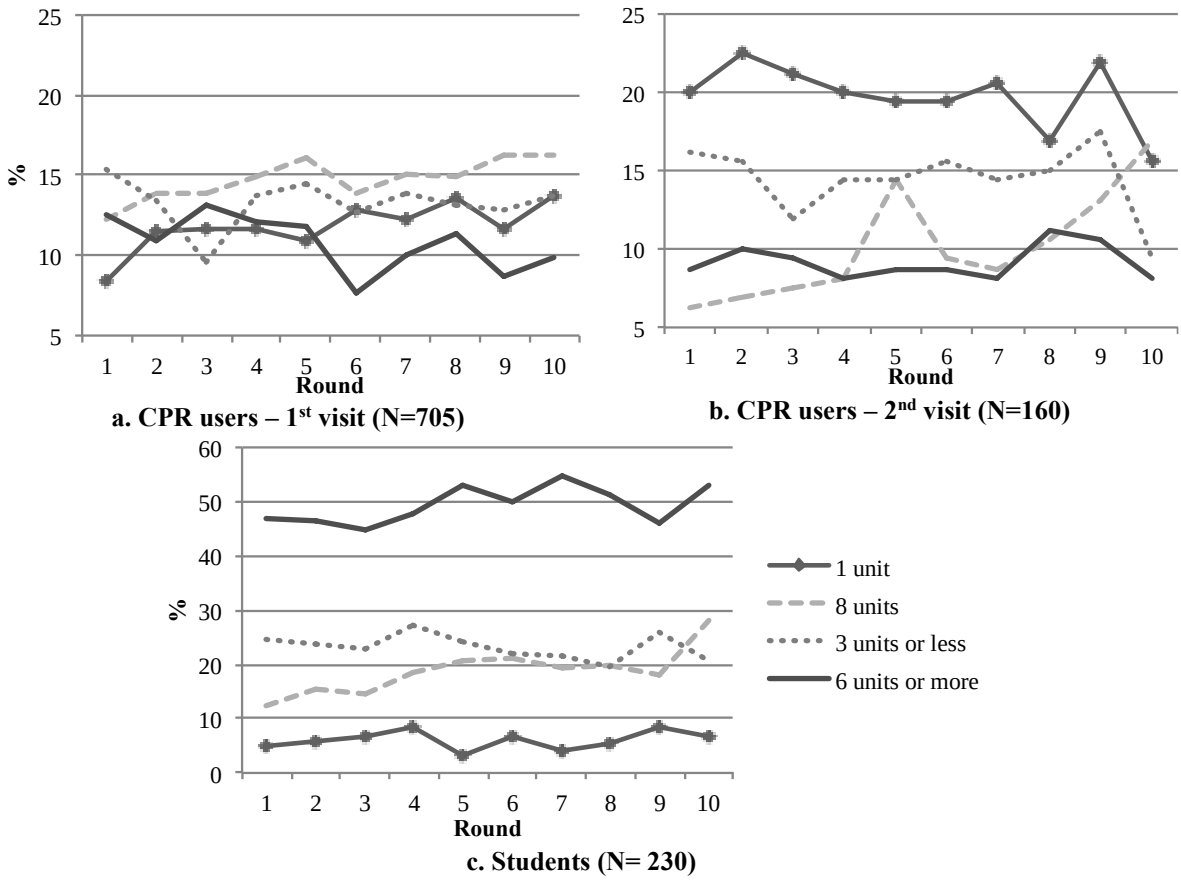


Figure B1. Baseline: behavior over rounds for Self-regarding, Pure Altruist and Percentage of participants that extracted 3 units or less, and 6 units or more. In panel a., the proportion of users that played the SNE is higher than the proportion of users that played the social optimum strategy. In panel b., in the second visit, the proportion of users who played the social optimum is the highest. In panel c., among students the proportion of individuals who played the social optimum is the lowest and the proportion of individuals who extracted 6 or more units of the resource is the highest.

Table B2. Real Users' perceptions on the CPR

Variable	Mean	Median	Standard deviation	Obs.
=1 if thinks that the CPR has not changed	22.0	0	41.44	678
=1 thinks the CPR is scarce	70.8	100	45.52	319
Perceived Percentage of families that cooperate in projects related to the CPR	56.6	50	38.84	708
Perceived interest in the community to cooperate	71.4	50	30.66	658
=1 if thinks the community should be the main vigilant	39.6	0	48.93	753
=1 if thinks the community should increase control of the CPR	52.6	25	36.54	743

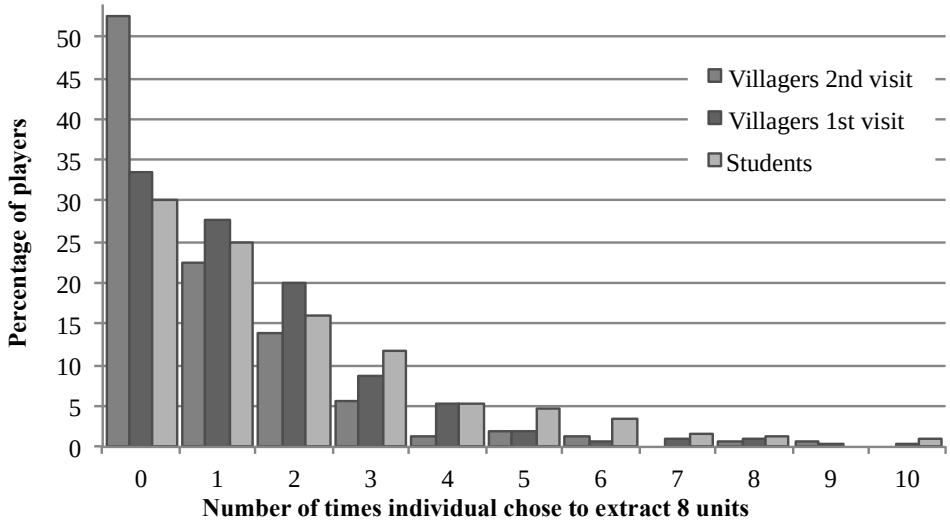


Figure B2. Comparison of self-regarding behavior -CPR users and students. Among students, 0.87% chose the Self-regarding NE strategy consistently during the first ten rounds. On the contrary, among CPR users in the first visit 0.28% chose the Self-regarding NE. The high percentage of 52.3 among CPR users that never chose to extract 8 units in the second visit is a signal of the difference in behavior from participants who knew the game with respect to those who played the game for the first time.

Appendix C. Individual best responses

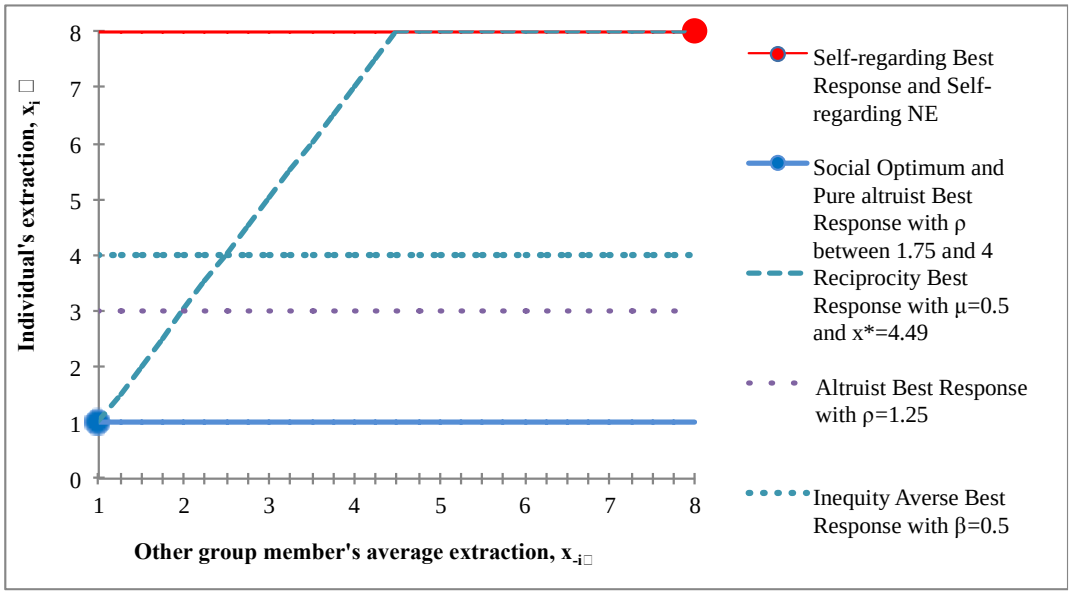


Figure C1. Best response functions example. Given the preference parameters $\rho_i = 0.08$ for an altruist, $\rho_i = 4$ for a pure altruist, $x_i^* = 4.49$ and $\mu_i = 0.5$ for a reciprocator, $\beta_i = 0.5$ for an Inequity averse.

Appendix D. Random Coefficient Model Classification

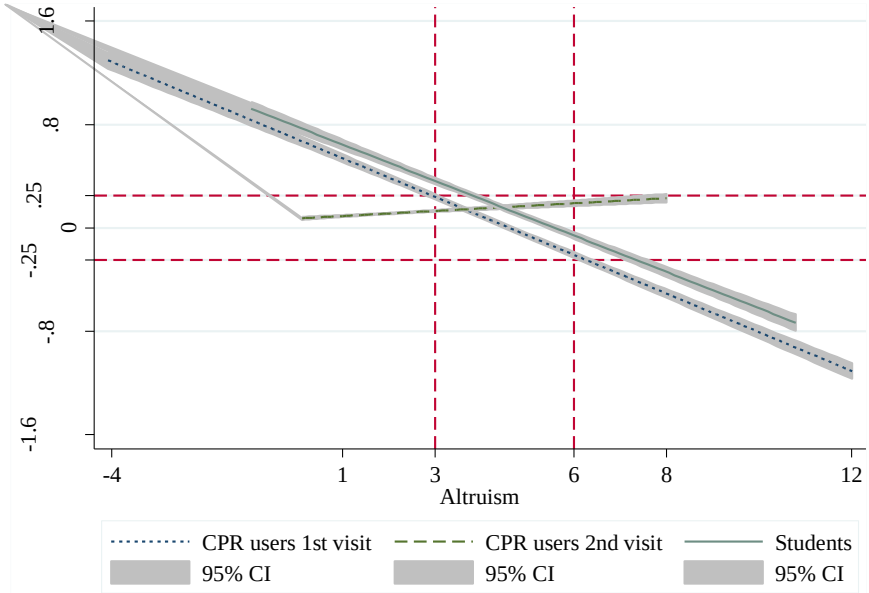
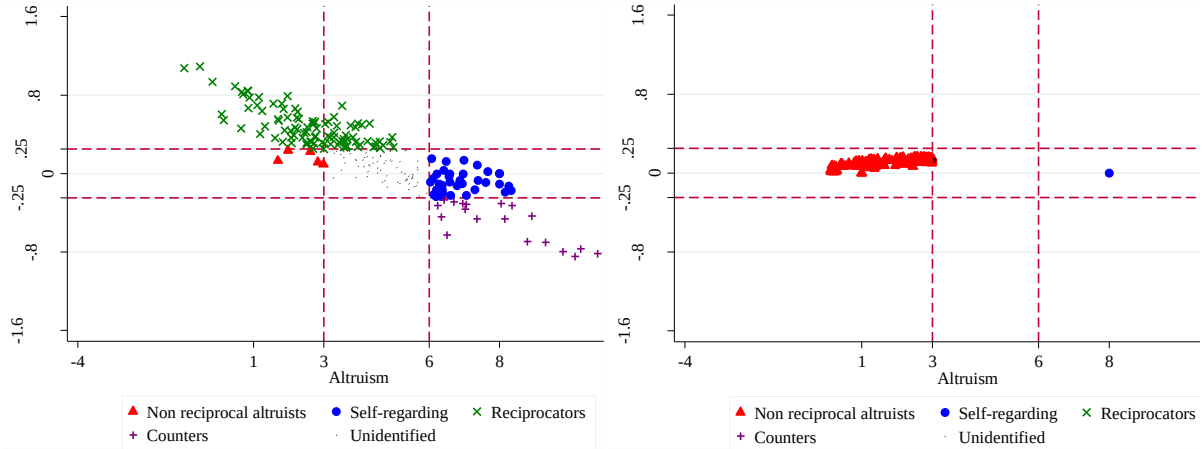


Figure D1. Linear Regression and confidence interval with 95% level of significance



a. Students (N=230) **b. CPR users – 2nd visit (N=160)**

Figure D2. The distribution of observed heterogeneity of altruism and reciprocity in the game with the RCM – Students and Second time visit CPR users

Table D1. Individual structural parameters estimators of Altruism and Reciprocity for each type – Second visit CPR users

Structural parameter		Non reciprocal Altruists	Reciprocator	Self-regarding	Counter	Unidentified
Altruism, $\hat{\rho}_i$	Mean	1.43	-	0.00	-	1.11
	Standard error	0.26	-	.	-	0.01
	Minimum	1.10	-	0.00	-	1.09
	Maximum	1.96	-	0.00	-	1.13
Reciprocity, $\hat{\mu}_i$	Mean	0.03	-	0.00	-	0.03
	Standard error	0.01	-	.	-	0.00
	Minimum	0.00	-	0.00	-	0.03
	Maximum	0.04	-	0.00	-	0.04
N		142	0	1	0	17

Notes: For CPR users in the second visit, we set $x_i^* = 3.83$, the average number of extracted units in the last practice round. The statistics for the Altruist type are very similar to CPR users in the first visit.

Table D2. Individual structural parameters estimators of Altruism and Reciprocity for each type – Students

Structural parameter		Non reciprocal Altruists	Reciprocator	Self-regarding	Counter	Unidentified
Altruism, $\hat{\rho}_i$	Mean	1.20	0.76	0.40	0.60	0.75
	Standard error	0.13	0.22	0.23	0.23	0.16
	Minimum	1.07	0.28	0.00	0.30	0.37
	Maximum	1.42	1.30	0.74	1.14	1.23
Reciprocity, $\hat{\mu}_i$	Mean	0.04	0.12	-0.02	-0.13	0.01
	Standard error	0.02	0.05	0.03	0.05	0.03
	Minimum	0.02	0.06	-0.06	-0.21	-0.06
	Maximum	0.06	0.27	0.04	-0.07	0.06
N		73	5	97	39	19

Notes: For Students, we set $x_i^* = 4.89$, the average number of extracted units in the last practice round. The statistics for the Altruist type are very similar to CPR users in the first visit.

Table D3. Social efficiency and group heterogeneity with the RCM- 2nd time visit CPR users

No. individuals in group g that are type q	Individual Type q label					
	Altruist	Reciprocator		Self-regarding	Counter	Unidentified
		$x_i^* > \bar{x}_{-i}^e$	$x_i^* \leq \bar{x}_{-i}^e$			
0	.	85.14	71.28	78.92	78.34	81.97
1	.	.	.	60.39	.	70.97
2	70.77	.	.	.	.	69.78
3	69.78	.	.	.	.	70.77
4	68.86	.	.	.	.	.
5	82.99	.	.	.	.	.

Number of groups is 32. The social norm is chosen as $x_i^* = \overline{x_{\iota 0}} = 3.84\ \forall i$, the average number of extracted units in the last practice round among students. *** represents significance at 1% level, ** at 5%, * at 10%.

Table D4. Social efficiency and group heterogeneity with the RCM- Students only

No. individuals in group g that are type q	Individual Type q label					
	Altruist	Reciprocator		Self-regarding	Counter	Unidentified
		$x_i^* > \bar{x}_{-i}^e$	$x_i^* \leq \bar{x}_{-i}^e$			
0	66.71	68.85	62.03	68.91	68.02	64.65
1	73.35	70.40	66.08	66.71	66.60	67.61
2	.	70.49	66.45	67.19	.	68.84
3	.	73.78	64.43	61.92	.	65.69
4	.	70.11	61.63	.	.	65.89
5	.	.	.	.	.	.

Number of groups is 46. *** represents significance at 1% level, ** at 5%, * at 10%. The social norm is chosen as $x_i^* = \overline{x_{\iota 0}} = 4.89\ \forall i$, the average number of extracted units in the last practice round among students.

Appendix E. Determinants of being type q

Table E1. Regressions for collinear variables

	Dependent variable: 1 if player is type	*Altruist*	*Self-regarding*	*Reciprocator*	*Counter*	$\hat{\rho}$	$\hat{\mu}$
	Independent Variables	I (5)	II (5)	III (5)	IV (5)	V (5)	VI (5)
=1 if one of two main economic activities in the household is	Agriculture	-0.045*	-0.021	0.041	0.002	0.879***	0.014*
		[0.027]	[0.020]	[0.038]	[0.035]	[0.023]	[0.007]
	Cattle	-0.036	-0.001	0.057	0.055	0.864***	0.008
		[0.032]	[0.027]	[0.057]	[0.049]	[0.035]	[0.013]
	Fishing	0.013	-0.064***	-0.001	0.019	0.911***	0.015**
		[0.027]	[0.017]	[0.035]	[0.037]	[0.024]	[0.007]
	Hunting	-0.035	-	0.142	-0.076	1.019***	0.028
		[0.067]	-	[0.116]	[0.099]	[0.086]	[0.028]
	Wood	0.085	0.034	-0.046	-0.033	0.928***	0.013
		[0.053]	[0.039]	[0.050]	[0.043]	[0.054]	[0.010]
Participant's perceptions	=1 thinks the CPR is scarce	-0.004	0.010	0.063	-0.026	0.888***	0.017*
		[0.036]	[0.024]	[0.053]	[0.052]	[0.028]	[0.008]
	Perceived Percentage of families that cooperate in projects related to the CPR	0.019	-0.010	-0.064	0.017	1.072***	0.003
		[0.032]	[0.020]	[0.051]	[0.041]	[0.029]	[0.007]
	=1 if thinks the community should increase control of the CPR	-0.082**	-0.024	0.083	0.018	1.060***	0.020***
		[0.038]	[0.026]	[0.051]	[0.048]	[0.034]	[0.007]
Social capital measures	=1 if membership in associations	-0.023	-0.045**	0.021	0.057*	0.899***	0.007
		[0.026]	[0.019]	[0.035]	[0.029]	[0.020]	[0.006]
	=1 if attends at least to one meeting	0.024	-0.045**	0.009	0.047	0.910***	0.009*
		[0.023]	[0.022]	[0.037]	[0.030]	[0.020]	[0.006]
	Number of associations meetings	0.000***	-0.002	0.000	0.000	0.005**	0.000
		[0.000]	[0.001]	[0.000]	[0.000]	[0.002]	[0.000]
	Number of days of volunteer work per year	0.000	-0.001	0.001	-0.001***	0.005***	0.000***
		[0.000]	[0.000]	[0.000]	[0.000]	[0.001]	[0.000]

* Significant at 10%; ** significant at 5%; *** significant at 1%. Robust standard errors in brackets clustered at the group level. For the types we report the marginal effects of a probit regression. For the parameters, we use a linear regression model.

Table E2. Determinants of being type *q*, additional controls (Table 10)

Dependent variable: 1 if player is type q		*Altruist*		*Self-regarding*		*Reciprocator*		*Counter*	
Independent Variables		I (1)	I (2)	II (1)	II (2)	III (1)	III (2)	IV (1)	IV (2)
Participant's characteristics	=1 if Landowner	0.018	0.009	-0.017	-0.011	0.055	0.060	-0.007	-0.001
		[0.024]	[0.017]	[0.024]	[0.020]	[0.066]	[0.067]	[0.054]	[0.052]
	=1 if the Household uses gas to cook	0.012	0.005	-0.002	-0.002	-0.036	-0.025	0.118**	0.125***
		[0.027]	[0.023]	[0.024]	[0.019]	[0.056]	[0.059]	[0.047]	[0.045]
	=1 if the Household uses electricity to cook	0.023	-0.022	0.080*	0.092	-0.038	-0.006	0.042	0.011
		[0.051]	[0.021]	[0.047]	[0.056]	[0.072]	[0.078]	[0.074]	[0.072]
Group composition (info within group)	% self-regarding among the rest of the group		0.073		-0.071		-0.336		0.217
			[0.088]		[0.065]		[0.241]		[0.193]
	Average age among the rest of the group		0.001		0.000		-0.001		-0.003
			[0.002]		[0.001]		[0.004]		[0.003]
Observations		379	379	379	379	379	379	379	379

* Significant at 10%; ** significant at 5%; *** significant at 1%. Marginal effects reported. Robust standard errors in brackets clustered at the group level.

Table E3. Determinants of being type *q*, additional controls (Table 11)

Dependent variable: 1 if player is type q		*Altruist*		*Self-regarding*		*Reciprocator*		*Counter*	
Independent Variables		I (1)	I (2)	II (1)	II (2)	III (1)	III (2)	IV (1)	IV (2)
Participant's characteristics	=1 if Landowner	0.018	0.009	-0.017	-0.011	0.055	0.060	-0.007	-0.001
		[0.024]	[0.017]	[0.024]	[0.020]	[0.066]	[0.067]	[0.054]	[0.052]
	=1 if the Household uses gas to cook	0.012	0.005	-0.002	-0.002	-0.036	-0.025	0.118**	0.125***
		[0.027]	[0.023]	[0.024]	[0.019]	[0.056]	[0.059]	[0.047]	[0.045]
	=1 if the Household uses electricity to cook	0.023	-0.022	0.080*	0.092	-0.038	-0.006	0.042	0.011
		[0.051]	[0.021]	[0.047]	[0.056]	[0.072]	[0.078]	[0.074]	[0.072]
Group composition (info within group)	% self-regarding among the rest of the group		0.080		-0.079		-0.291		0.190
			[0.097]		[0.061]		[0.237]		[0.194]
	Average age among the rest of the group		0.001		0.000		-0.001		-0.003
			[0.002]		[0.001]		[0.004]		[0.003]
	Average Index of time spent by the HH on CPR extraction among the rest of the group		0.098		-0.111		0.102		-0.187
			[0.068]		[0.072]		[0.175]		[0.193]
Observations		379	379	379	379	379	379	379	379

* Significant at 10%; ** significant at 5%; *** significant at 1%. Marginal effects reported. Robust standard errors in brackets clustered at the group level.

Table E4. Linear estimation of determinants of parameters $\hat{\rho}$ and $\hat{\mu}$. additional controls (Table 11). First visit CPR users.

	Dependent variable	$\hat{\rho}$				$\hat{\mu}$			
	Independent Variables	V (1)	V (2)	V (3)	V (4)	VI (1)	VI (2)	VI (3)	VI (4)
Participant's characteristics	Sex (=1 if woman)	0.035	-0.017	0.041	-0.024	0.017	0.007	0.017	0.005
		[0.036]	[0.037]	[0.037]	[0.038]	[0.012]	[0.012]	[0.012]	[0.012]
	=1 if Landowner`	0.078	0.008	0.073	0.013	0.023	0.015	0.023	0.015
		[0.049]	[0.042]	[0.050]	[0.042]	[0.016]	[0.016]	[0.016]	[0.016]
	=1 if the Household uses gas to cook	0.162***	0.076*	0.176***	0.080**	-0.021*	-0.024**	-0.021*	-0.021*
		[0.044]	[0.039]	[0.047]	[0.040]	[0.012]	[0.012]	[0.012]	[0.013]
	=1 if the Household uses electricity to cook	-0.013	-0.050	-0.007	-0.049	-0.008	0.001	-0.009	-0.001
		[0.066]	[0.063]	[0.064]	[0.061]	[0.016]	[0.016]	[0.016]	[0.016]
Participant's perceptions	=1 if thinks the community should be the main vigilant	0.016	0.016	0.019	0.023	0.007	0.008	0.007	0.008
		[0.039]	[0.032]	[0.039]	[0.033]	[0.010]	[0.010]	[0.010]	[0.010]
Group composition (info within group)	Average $\hat{\mu}$ among the rest of the group		0.058		-0.013		0.157		0.171
			[0.224]		[0.224]		[0.140]		[0.144]
	Average sex among the rest of the group		0.073		0.076		0.020		0.022
			[0.050]		[0.051]		[0.018]		[0.018]
Observations		379	379	379	379	379	379	379	379
R-squared		0.873	0.900	0.872	0.901	0.058	0.088	0.058	0.089

* Significant at 10%; ** significant at 5%; *** significant at 1%. Robust standard errors in brackets clustered at the group level.

III. Building social capital: conditional cash transfers and cooperation[35]

This chapter is joint work with Orazio Attanasio♣ and Luca Pellerano♦. Forthcoming in the Journal of Economic Behavior and Organization.

Abstract

Many conditional cash transfer (CCT) programs have important social components and, therefore, can have an effect on social capital. In 2007, we conducted a field experiment with 1451 subjects in Cartagena, Colombia. We interpret the behavior in the game as a measure of what in the literature has been called social capital. We played the game in two similar and adjacent neighborhoods. The 'treatment' neighborhood, *Pozón*, had been targeted for over 2 years by a CCT program, Familias en Acción; the 'control' neighborhood, *Ciénaga*, had not. In 2008, with the program being implemented in both neighborhoods, we played the same public goods game, and were therefore able to implement a difference in differences strategy to estimate the impact of the CCT on our measure of social capital.

In 2007, the level of cooperation we observed in the treatment neighborhood was considerably higher than that in the control one. Although similar in many dimensions, the two groups turned out to be significantly different in some observable variables; the positive result was robust to controls for these differences. In 2008, we found that the level of cooperation was statistically identical across the two neighborhoods, and similar to the levels observed in 2007 in the treatment one. We conclude that the CCT program did improve cooperation.

In analyzing the effect of the CCT on cooperation we also look at other (individual and group) determinants of individual behavior in the game, and we compare our measure based on behavior in the game to more traditional measures of social capital used in the literature that we collected in a context-specific survey.

JEL Codes: C92 (Experiments Laboratory, Group Behavior), D70 (Analysis of Collective Decision-Making); D78 (Policy making and implementation); H41 (Public goods); Z13 (Social norms and social capital)

Keywords: Behavioral experiments, social preferences, social capital, conditional cash transfer programs, cooperation, public goods provision

[35] We thank Liliam Puello and Beatriz Jiménez from the local office of Familias en Acción in Cartagena, and Hernando Sánchez and Ana Tamayo from the national office of Familias en Acción. Without the logistic support of the program officials at all levels this exercise would not have been possible. We also thank Syngjoo Choi, David Echeverry, Patricia Padilla and Vivian Rodríguez. The experiments reported here were sponsored by the Corporación Andina de Fomento -CAF-, the European Commission and Institute for Fiscal Studies and EDePo. Attanasio's research was partly financed by European Research Council Advanced Grant 249612 on "Exiting Long Run Poverty: The determinants of asset accumulation in developing countries" and by the ESRC/DfID grant "Understanding external determinants of the effectiveness of Cash Conditional Transfers: a benchmarking investigation" ES/H034374/1. Liliam Puello passed away in November 2011. We dedicate this paper to her.

♣ University College London, IFS, NBER, CEPR and BREAD.

♦ IFS and Oxford Policy Management

1. Introduction

Social policies may also improve economic outcomes through changes in the structures of social relationships and their forms of organization. Social capital refers to the set of resources that inheres in those relationships and the structure of such relationships (Uphoff, 1999). It can be understood in terms of social norms and networks (Putnam, 1993, Coleman, 1990) and it manifests itself in patterns of pro-social behavior (trust, reciprocity, and cooperation (Christoforou et al., 2014).

Conditional Cash Transfers (CCTs) have become one of the most popular interventions in developing countries. There is a strong line of research on the effects of CCTs that show that CCTs are successful in their goals as social assistance, raising service use and improvements in target outcomes (Fiszbein et al., 2009, Adato and Hoddinott, 2010). Although CCTs' designs were meant to have a 'social capital' approach[36] and anecdotal and qualitative evidence has been found (Cohen and Franco, 2006), rigorous studies on whether CCTs affect social capital has been a prevalent interest just recently and mostly on voting behavior[37]. Although the main goals of the CCTs are not related directly to improve participation and collective action, the fact that most of CCTs have a component of new interactions among the beneficiaries within the communities would lead to build or strength social ties (Putnam, 1995) and enforce pro-social norms.

In Colombia, there is a large CCT program, called Familias en Acción (FA) that has, like many other similar programs, especially in Latin America, a strong social component. Beneficiaries meet periodically and elect a representative that coordinates a number of activities in which they are involved. It is reasonable to argue that these activities could increase cooperation, trust and, more generally, interactions among beneficiaries that could improve the social fabric of poor communities and, potentially, improve collective action and the ability to overcome social dilemmas. The main goal of this paper is to test this hypothesis exploiting quasi-experimental variation induced by the roll out of FA to urban areas.

In this study, we estimate the impact of the introduction of a conditional cash transfer program on social capital. The first challenge we had to face was to devise a measure of social capital in the context we were studying. On one hand, integrating quantitative with qualitative measures as an improved approach to measure social capital in program evaluation is not new (Adato, 2008, Casey et al. 2012) but evidence is not conclusive[38]. Adato (2000) uses qualitative

[36] For example, Juntos (Together) in Peru, Puente (Bridge) in Chile, FA (Families in Action) in Colombia, Progresa (Progress) in Mexico, Bolsa Familia (Family pot) refer all to features of social capital: networks, 'bridging' social capital, family social capital, 'linking' social capital, cooperation (ECLAC, 2005, Arriagada and Mathivet, 2007). ECLAC (2005) present this approach in detail.

[37] Some recent evidence on the positive political impacts and voting behavior of CCT programs in Colombia (Nupia, 2012) and Zárate et al., 2013), Brazil (Zucco, 2011), Romania (Pop-Eleches and Pop-Eleches, 2012), Uruguay (Manacorda et al., 2011), Philippines (Labonne, 2013) and Mexico (De la O, 2013)

[38] There is a strong line of research on the impact of ''community-driven development'' (CDD) programs whose main goal is to actually build social capital and seem to have no effect on the desired outcomes. Although, many of these studies suggest the existence of bias in self-reported levels of pro-sociality and social cohesion from surveys (Avdeenko and Gilligan, 2014, Mansuri and Rao 2013, Wong, 2012, Fearon et al. 2009, Casey et al. 2012), very few conduct labs in the field (Avdeenko and Gilligan, 2014, Fearon et al., 2009, Nguyen and Rieger, 2014, Alzua et al. 2014). Fearon et al. (2009) employed a framed public goods game and found a positive impact of a CDD initiative on individuals' ability to overcome collective action problems in Liberia. Nguyen and Rieger (2014) found a positive impact on a CDD and contribution in a public goods game in Morocco.

data (e.g. focus groups and interviews) and finds positive link between social capital and the CCT Progresa in Mexico whereas Camacho (2014) uses survey measures only and finds no effects of the local CCT on social engagement in Peru[39].

On the other hand, despite the complexity of conceptualization of social capital, what lies at its heart is the ability to overcome free-rider incentives in real-world situations (Coleman, 1988). These situations, that especially poor communities face, share the same game-theoretic representation. Then, as a community level attribute (Coleman, 1990), in order to build social capital, individual behavior might be subject to "free-riding" (Samuelson, 1954; Olson, 1965; Grossman and Hart, 1980; Ostrom, 1998). Evidence on social programs evaluation supports this claim (Adato et al., 2005, Fearon et al., 2009, Avdeenko and Gilligan, 2014, Alzua et al, 2014).

For this reason, we propose a measure of social capital based on the behavior in a public goods game. The use of 'lab in the field' experiments as a method to study social preferences within a community and to measure social capital is not new[40]. In particular, public good games have been used in a variety of different situations, both urban and rural, and are one of the experimental methods that have been often used in the literature to examine behavioral motivations such cooperation in groups. Our design has been used extensively in Colombia (Cárdenas and Jaramillo, 2007, Cardenas et al. 2013, Attanasio and Pellerano 2012b) and in other countries (Cardenas et al. 2013, Barr et al., 2014). Our original contribution is to implement the game to measure the impact of a specific intervention in a quasi-experimental setting.

The Public Goods game we use to measure social capital captures the willingness to cooperate among the members of a group of 25 people by choosing simultaneously whether to allocate a token in the private account with a private benefit or to allocate the token in the group account, where the benefits of all members increases and the well-being of the entire group is improved. The sum of tokens allocated to the public good is then multiplied by a factor to model the additional value generated from the public nature of the good and then shared equally among the group members. The situation just described constitutes a typical social dilemma. There is no incentive to invest in the group account due to a higher individual payoff by investing in the private account. The dominant strategy is not to contribute at all, undermining the social outcome. However, if all in the group invest their token in the private account (assuming players maximize their own monetary payoff and rationality is common knowledge), the group will be worse-off than if all the members invest in the group account, which constitutes the social optimum. The experimental literature has extensively documented that, in similar contexts, the Nash equilibrium is not observed, either in the lab or in the field. Groups of individuals seem to be able to internalize at least in part the externality built in the game[41]. For this reason, this set up seems particularly attractive in devising a measure of 'social capital'.

The possibility of cooperation within a group is determined by multiple factors such as

[39] There is similar evidence from Cash Transfer (CT) programs (MacAuslan and Riemenschneider, 2011, in Malawi and Zimbabwe and Cameron and Shah, 2014, in India). However, Pellerano et al, (2014) find a positive link of a CT and engagement in reciprocal sharing activities in the community in Lesotho.

[40] Carpenter (2002), Carpenter, et al. (2004), Gaechter et al (2004), Karlan (2005), Cárdenas et al. (2009), Fearon et al. (2009), Voors et al. (2012), Gilligan et al. (2014), Jakiela et al (2014).

[41] In general, contributions to the public good are in a range of 40%-60% of the group optimum (Ledyard, 1995, Croson 1996, Camerer and Fehr, 2004, Attanasio et al., 2009). At the same time, there seems to be considerable amounts of heterogeneity in the ability different groups have to solve this kind of problems.

repetition, communication, punishments or rewards and inequality in the payments. In our game, the incentives to invest in the group account are given by the specific features of the design, but also by the individual motivations concerning the group wellbeing. Individual attributes such as altruism, trust, social distance from the other members (Cárdenas, 2003), fairness (Rabin, 1993), reciprocity (Andreoni, 1988, 1995, Bowles and Gintis, 2004), a sense of affiliation as a member of a common group, or sympathy toward others in the group (Attanasio et al., 2009) determine social cohesion in a group and strengthen the ability of its members to cooperate and overcome collective action problems. In addition, community attributes that would be affected by the CCT, such as social norms and institutions, informal enforcement mechanisms, concerns for social reputation, social reciprocity (Bowles and Gintis, 2004) and group identification enforce the group interests over the individual, leading to attain a higher level of contribution and overcome the dilemma. We expect that our design will allow us to examine how cooperation has been affected by the presence of a CCT.

The first challenge we had to face is related to how we estimate the effect of the CCT on willingness to cooperate. To evaluate the impact of an intervention one would like to rely on a Randomized Control Trial. As the outcome we are interested on is the 'social capital' in a community, one would ideally randomize the intervention across a number of communities. Unfortunately, such a design was not possible. We therefore rely on evidence on individual behavior and comparing the choices of individuals living in a community with the program with that of individuals living in a community without the program. Of course, any observed difference between the 'treatment' and 'control' individuals could be driven by systematic differences at the community level in the level of social capital that are not related to the program. To circumvent this problem, therefore, we rely on a relatively standard diff in diff approach and estimate differences between the two communities we consider that are unrelated to the program by observing behavior in a period in which both communities are either both exposed or both not exposed to the program.[42] For such a strategy to be valid it is necessary that systematic differences between the two communities are constant over the period considered and determine behavior in an additive fashion. In the literature that uses diff in diff this assumption is often referred to as a common trend assumption. In what follows, we discuss its plausibility.

2. Familias en Acción and its expansion

Familias en Acción is a CCT that was inspired by the Mexican CCT PROGRESA and whose goal was to reduce extreme poverty in the medium term by providing resources to improve the nutritional status of poor households and in particular their children as well as school enrolment. To get access to the program's grants, beneficiary households have to comply with a number of requirements. FA has three components: a nutritional and health component aimed at households with children less than five, an education grant for children in primary school and an education grant for children in secondary school. The health and nutrition grant, roughly equal to US$25 independent of family size, is conditioned on attending regularly growth and development check-ups for children, a vaccination program and some 'classes' on hygiene, diet and contraception. The educational grants, aimed at households with children aged seven to seventeen, are conditional on enrolment and regular attendance in

[42] In our case we have one year (2007) when one community was exposed to the program and the other not and one year (2008) in which they were both exposed to the program. We therefore use 2008 to identify the systematic differences between the two communities and subtract that from the difference observed in 2007 to obtain the effect of the program.

school. Each child in primary (secondary) school entitles the household to about US$8 (US$16) per month. Households receive a total transfer which may oscillates between 10% and 21% of the minimum wage and between 25% and 50% of the average level income of the poor (DNP, 2010 and MESEP, 2012).

The program also includes an important social component, articulated around periodic meetings of beneficiaries, called *Encuentros de Cuidado* (EC) [Care follow-up meetings]. Although participation to these meetings is neither compulsory nor part of the conditions to receive the transfer, most beneficiaries (95.94% in our sample) participate in the EC where, in addition to discussing hygiene, nutrition or other health-specific issues, they have the possibility to talk about different topics or even simply chat. Beneficiaries are invited to attend the meetings, which are introduced as key for human capital investment. Conversations with program's officials and with beneficiary mothers indicate that these social aspects are indeed an important feature of the program: beneficiary mothers start new activities, get to know each other better and improve their ability to act as a group.

Additionally, the beneficiaries elect a representative, called *Madre Líder* (ML) who is in charge of communication with the local office and is also in charge of organizing the social activities and educational meetings (such as the EC) to which the beneficiaries are supposed to participate. As a consequence, the ML's often assume a prominent and visible role in the community[43].

As the theory of social capital suggests, the fact that our CCT would require people to interact with others may create an environment for them to experience the potential of social networks to improve their wellbeing. Our CCT would affect social capital once the program is able to affect social interactions and their environment (Coleman, 1988). FA may create networks or strengthen the current ones and improve the structure of social relationships among beneficiaries (Putnam, 1995), promote leadership (Latham and Saari, 1979, Bass, 1991) and give mothers the opportunity to start working as a "social group" by perceiving a strong identification with the program (Tajfel, 1982) and their power to act 'together' (Warren, 1998). This facilitates group decision-making and increases the willingness to intervene for the common good, which eventually could lead to overcoming social dilemmas by enforcing pro-social norms (Coleman, 1990).

Empirically, we are able to distinguish between attitudinal (i.e. perception of cooperation or trust in the community) and behavioral social capital (i.e. voting behavior, participation in meetings and membership in associations as well as willingness to cooperate, social networks and social norms) (Carpenter, 2002). Our CCT may increase attitudinal social capital in some ways. First, by becoming a beneficiary a new group identity emerges which would change the perception of the community traits. Beneficiaries share the same paperwork load, health check-ups, payment logistics and the same interests. Second, the EC and beneficiaries' assemblies are a place of encounter with people that face the same needs and interests. This continued interaction among beneficiaries would create and enforce social norms so beneficiaries' perception of trust is also modified.

On the other hand, our CCT may increase behavioral social capital, first, by transforming voting behavior as confidence/perceptions on the current government or a specific political party change. Second, when social relations among members of a community are strengthened,

[43] By 2007, there were 800 ML with group size ranging between 50 to 80 beneficiaries.

incentives to participate in organizations or meetings in order to solve problems within the community may be stronger as well.

The program has become the flagship of the Colombian government's social policy as it targets the poorest 20% of Colombian households[44]. It started in 2002 in 627 small rural areas and in 2007 was expanded to all urban areas in order to include 1.5 million beneficiary households[45]. In 2005, the FA authorities decided to pilot the program in *Pozón*, a neighborhood in the city of Cartagena. Between 2005 and the first half of 2007, the program operated in *Pozón* but had not been implemented in other neighborhoods, despite there being other two neighborhoods (*Nelson Mandela* and *Ciénaga de la Vírgen)* identified by the FA authorities as eligible to participate in the pilot[46].

In late August 2007, just a month after our first game, a new wave of massive enrolment to the program started in every municipality in the country, regardless of its population. The program was also rolled out in all the poorest neighborhoods of Cartagena, including *Ciénaga*.

3. Data

In this paper, we utilize two types of data. First, we analyze the behavior of individuals in the field experiment that we describe below. Second, we use extensively the survey we collected at the end of the experiment. In what follows we describe the recruitment of subjects for the game in the two neighborhoods, the protocol of the games and the data we collected.

Table 1. Number of players and sessions

			Individuals			Sessions [c]		
			Pozón	**Ciénaga**	**Total**	**Pozón**	**Ciénaga**	**Total**
Baseline	**2007**		342	334	676	14	14	28
Follow-up	**2008**	Panel	237	290	527	7	11	18[b]
		Independent cross section	404	371	775	14	15	29[a]

[a] Sessions with New recruits, i.e. with 95% or more people who were playing the game for the first time (i.e. 714 subjects). [b] Sessions with Old recruits, i.e. with 95% or more people who were playing the game for the second time (i.e. 443 subjects). [c] Sessions that were neither New nor Old (at 95% composition threshold) are not in the Table. There were 5 of them in Pozón and 1 in Ciénaga, corresponding to 145 subjects.

The data for the entire set of experimental and survey data contains information on a sample of 676 and 1302 participants in the baseline and the follow up, respectively. In 2007 we conducted 28 sessions, 14 in each neighborhood. In 2008 we conducted 47 clean sessions, 21 in *Pozón* and 26 in *Ciénaga*[47], which constitutes 775 new participants and 527 individuals that had

[44] In Colombia, most welfare programs are targeted using the so-called SISBEN score, a poverty indicator that is updated periodically. On the basis of this score, households are assigned to one of six categories. FA targets the level 1 of SISBEN and displaced people.

[45] For evidence of success of FA on the target outcomes see Attanasio et al. (2005), Attanasio et al. (2010), Attanasio et al. (2012a), Attanasio et al. (2012b), Baez and Camacho (2011); for evidence about other outcomes such as crime and voting behavior see Camacho and Mejia (2013) and Zárate et al. (2013). Attanasio et al. (2012b) shows that these positive effects are still noticeable amongst beneficiaries 10 years after their first enrolment in the program.

[46] Ciénaga and Pozón were considered very similar to the local authorities and are adjacent. The third neighborhood (*Nelson Mandela)* is geographically distant and has received a much larger influx of displaced households from the countryside. For more information see the appendix A in the appendix.

[47] In 2008 we conducted 53 sessions, 26 in *Pozón* and 27 in *Ciénaga.* However, the complexity of the logistics led to 6 sessions in which more than 5% or less than 95% participants were new or former participants (called mixed sessions). We exclude them from the analysis.

played in 2007[48]. These figures are reported in Table 1. Overall we obtained a sample with 1,451 of individuals that did not participate in any game before.

i. Experimental set up

In our design, a Voluntary Contribution mechanism (VCM), each individual in a session of 25 participants is given one token. The player has a simple discrete choice to make: she can either invest the token in what we label a 'group' account or in a 'private' account. Tokens placed in private accounts are eventually redeemed at $5[49], while those placed in the group account are redeemed at $10. Players who placed their tokens in the private account will be entitled to $5 and an equal share of what is in the group account. The amount in the group account is divided among all players, regardless of who contributed to it: given that there are 25 players, each would receive $0.40 for each token invested in the group account by any other member in the group. Individuals' private decisions are made simultaneously and contributions are private and anonymous, that is, players are not allowed to communicate. Those who put their tokens in the group account are only entitled to a share of what is in the group account. Note that a player contributing to the group account will get more than the $5, which is the minimum she gets if she contributes to the private account, if at least 13 players contribute to the group account[50].

The marginal propensity to cooperate also called marginal per capita return (MPCR) in this game is one of the lowest in the literature. The goal was to mimic a measure of what is called bridging social capital, the ability to overcome social dilemmas in a very large group. Instead, we argue it matches more closely the reality we are trying to depict. Given the level of deprivation in the neighborhoods we study, the intensity of the social dilemma is arguably much higher than in most other lab, or even lab-in-the-field, studies, and a low MPCR is better fitted than a higher one.

More importantly, the low MPCR makes our results more forceful. Because the power of our analysis would have been maximized if we had had a MPCR of 50%, a low value provides evidence that if the coefficient is subject to any bias it will be downward bias. This is particularly important in light of our main goal. The abovementioned studies do not measure social capital but rather correlations between game outcomes and decisions in other contexts. Because we claim to provide a social capital measure[51], where social dilemmas are key -and hence a low MPCR appropriate-, the low MPCR provides an additional source of validity to the measure.

The dichotomous VCM makes the game easily understood by subjects and also time effective.

---

[48] In 2008 we were faced with the issue of whether to play with the same individuals or to recruit a fresh sample of players. Both strategies had advantages and disadvantages. Playing with 'former' players would allow us to construct a longitudinal data set and potentially control for individual fixed effects. However, behavior in 2008 would be conditioned to having experienced the game within a specific group in the previous year and it would require a larger sample and more funding to take this effect into account.

[49] Units in thousands; thus $5 means COL$5000 (US$2.85, according to the official exchange rate at that date TRM: US$1=COL$1753.01 (monthly mean average for July 2008, http://www.oanda.com)).

[50] At the end of the session, only the percentage of players contributing to the group account is revealed and players are paid after they take a short socio-economic survey. Participants played the game for 2 rounds. After playing the first round, individuals are allowed to talk and communicate. Ten minutes later, there is a second round. For more details, see Appendix.

[51] In fact we claim to have a measure of the effect of the program on a measure, as opposed to a mere correlation, which arguably makes the point more relevant.

This is perhaps one of the reasons why only two other studies employ a dichotomous VCM game comparable we employed in Colombia: Cardenas et al. (2013) in 6 Latin American cities, Alzua et al. (2014) and Barr et al. (2014) in Albania.

In July 2007 and July 2008 we conducted the Public Goods game described above in *Pozón* and *Ciénaga*. The VCM game we described above is played in two rounds. The subjects first decided whether to contribute to the group account individually and privately, without having the possibility of communicating with other players. After the first round (but before its results are revealed), they are given the possibility of talking about whatever they like. Then subjects had to decide again what to do with their token, individually and privately, before the results of the first round are announced. The role of communication and beliefs on individual willingness to cooperate is far beyond our research question. In this paper we only analyze willingness to cooperate in the first round.

The protocol followed in the two years was identical. Participants were invited to come to the local public school. After collecting their identification documents and checking their names on the recruitment lists, subjects in each session were given an identification number randomly and seated in semi-circle in a classroom where the instructions of the game were read and explained. After the participants played the second round of the game described above, we collected a network questionnaire on the existing relationships among them and a relatively short survey that gathers information on a wide range of socio-economic variables.

All recruited people were given US$1.2 as part of their show-up fee in order to induce credibility and to subsidize the transportation cost from their homes or workplace to the school we assigned for the experiments stage. Once the session ended participants were paid their earnings based on the decisions in the experiments. On average each participant earned US$9.6 (COL$16867)[52].

ii. Sampling, recruitment and allocation into sessions

We recruited the participants to the game with the help of the local office of FA (Enlace Municipal). For the first stage of our study, in 2007 the recruitment in the treatment and control neighborhoods was slightly different. In Pozón, where the program was already operating, we could contact directly beneficiaries. The ML representatives sent an open invitation to all beneficiary mothers. In Ciénaga, instead, beneficiary mothers had not yet been registered into the program. However, since the program was already planning the expansion in, the office also had list of 'future' beneficiaries in that neighborhood. These lists were used in recruiting the participants in Ciénaga. In addition, we also used an open communication strategy: a person with a loud speaker announced the invitation twice a day, one week before fieldwork started.[53]

In both localities, mothers were allocated into session as a first-come first served basis and there was no control over who attended which session, only whether the sessions were treatment or control and the participants dwelled in their respective neighborhood. The FA list and the open communication campaign in Ciénaga proved to be effective in identifying future FA beneficiaries, as 92.51% of participants in these sessions were subsequently selected by FA as beneficiary households in 2007[54].

[52] FX rate US$1=COL$1,951.84 (monthly mean average for July 2007, http://www.banrep.gov.co)
[53] This is a method commonly used to send messages in the slums, given the lack of telephones.
[54] In particular, 82.34 % of participants in these sessions were subsequently selected by FA as beneficiaries in late 2007 and 10.18% of participants belonged to a household with a FA beneficiary.

Unlike in 2007, in 2008 when we conducted the game for the second time, we could use the same recruitment strategy in both neighborhoods, as beneficiary lists were available in both neighborhoods. Besides contacting old participants, invitations were also sent to 500 randomly selected new participants from the FA beneficiaries list in each neighborhood[55]. The FA office sent the invitations through the ML to those specific households in order to attend to any of the sessions held (a span of 3 days).

Conducting lab in the field experiments in large cities provide many challenges to the experimental design in terms of costs, time, recruitment and attendance rates (Ñopo et al., 2008, Candelo and Polanía-Reyes, 2008) For example, implementing sessions with 25 randomly allocated individuals was impractical and infeasible. The two neighborhoods are a 2 hour-drive apart; in order to minimize 'cross-talk' and its effects – participants talking about the experiment to future players who will participate in subsequent sessions, sessions were implemented in a four-day frame with four sessions each day in each neighborhood. For example, during the first four days we conducted the experiments with new participants in Pozón and the following four days with new participants in Ciénaga. Since the time frame was too short, we gave the beneficiaries the freedom to choose the session that bests suits them. The requirements were to be in the list of randomly chosen beneficiaries. It could have been the case that people who might be neighbors arrived together to the sessions. In fact, some invited beneficiaries arrived to the session in groups[56].

There is the possibility of contamination among subjects of different sessions: participants to a session could talk to participants to the next session on the way out, although we put lot of effort in avoiding these contacts. This type of contamination would be the same or very similar in both treatment and control. In order to control for this possibility we included in our analysis the average outcomes in previous session and the number of sessions that had occurred before.

iii. The post-game survey

In the survey we collected at the end of the session, there is a module on individual socio-economic characteristics (e.g. age and level of education), a module on household characteristics (e.g. income, assets, household size), a module on social capital measures (e.g. membership in associations, trust), a module about FA (e.g. date of enrolment), a module about the game (e.g. whether the participant understood the game).

iv. Measuring social capital in a FA locality

The data we collected both during and after the games allowed us to construct several social capital measures. In what follows, we consider three different sets of measures: those derived from the behavior we observe in the games, those we can derive from the socio-economic questionnaire and those we derive from the network data we also collect.

[55] Old participants were sent an invitation and were allocated in a session with a specific day and time in order to reproduce the same sessions we hold in 2007.

[56] The fact that individuals are not randomly allocated into sessions allowed us to explore the role of social networks in the effect of the program on cooperation. We were able to obtain enough variation in terms of the density and quality of the network across the neighborhoods and compared the results with those of pre-existing networks in Ciénaga 2007. We also performed a sensitivity analysis and included the network data in the regression analysis, which did not affect the results.

First, from behavior in the game, we consider the propensity to contribute to the public good as a measure of attitude to cooperate in a group. An index of social capital could then be the proportion of individuals who contribute to the public good. This measure captures the ability of a group to internalize an externality and the willingness individual players have to act as a group.

Second, although the use of direct survey questions on trust has been strongly questioned as a poor indicator of social capital in comparison with experimental outcomes, in order to compare our data with the literature, we included in the postgame survey traditional measures of social capital regarding trust (such as those derived from the WVS), membership in organizations and participation in political and social processes within the community. We also considered perception of cooperative attitudes. These variables have been widely used as a proxy for social capital in the non-experimental literature.

Finally, in our survey, each player was asked about her relation with all the other players, where the options given were: (a) relative, (b) friend, (c) acquaintance or (d) unknown. In addition, we also asked questions about who would the participant consider as a leader in the community. For every session, we are able to construct a relationship matrix that describes the shape of existing networks among players. This information allows us to calculate a set of variables on session connectivity and eventually control for the strength of relationships in each group. We construct measures of how well connected the participants of each session were. The simplest measure of the "social integration" of each player within the session is the number of acquaintances, relatives and friends a participant had in her session. We also use an index of connectivity defined as the total number of connections (i.e. as an acquaintance, as a friend or as a relative) in the session and a dummy for whether she was considered a leader (See appendix A).

Despite social networks being a key dimension of social capital to improve socioeconomic outcomes (Coleman, 1990, Putnam, 1995), we don't focus on our networks measures since individuals were not randomly allocated in the sessions. Instead, we focus on our behavioral measure of social capital and we do describe the other measures for illustrative purposes.

v. Descriptive statistics

Attanasio et al. (2009) reports descriptive statistics and the tests of significance for the difference between the two neighborhoods in 2007. These statistics are also reported in the first panel of Table 2 and give a good representation of the main characteristics of this population. Participants come from very poor families, with low levels of income and education. Although the two neighborhoods are very similar in many dimensions, and they are next to each other, important differences emerge quite clearly. Participants from *Pozón* were significantly less likely to be head of household (indicating a larger presence of partners) and have better access to public services. Moreover they are more likely to be educated. Most importantly, participants from *Pozón* seem considerably less poor than those in *Ciénaga.* They are more likely to own the house where they live, to be connected to piped water, to own several durables and other assets and less likely to have a dirt floor in the house (although there is no significant difference in the level of income or food security). They are also less likely to be unemployed. Some of the differences observed in 2007 could have been induced by the program and, given the nature of the data; it is hard to disentangle the effect that FA may have had on many of the socio economic outcomes in both neighborhoods. For instance asset tenure and monthly income may be clearly affected by the fact that participants

from *Pozón* had already been received the benefits from FA.

Table 2. Demographic characteristics of the participants - 2007 and 2008

		2007			2008		
	Variable	**Pozón**	**Ciénaga**	**Difference**	**Pozón**	**Ciénaga**	**Difference**
General characteristics	Percentage of female participants	100	97.9	2.1**	98.8	98.1	0.7
		(0.0)	(1.1)	(1.0)	(0.7)	(0.7)	(0.9)
	Average age (years)	38.0	33.6	4.5***	36.2	36.2	0.0
		(0.6)	(0.7)	(0.9)	(0.6)	(0.4)	(0.7)
	Number of years living in the neighborhood	14.5	22.4	-7.9***	14.6	22.3	-7.7***
		(0.3)	(0.5)	(0.6)	(0.5)	(0.6)	(0.7)
	Percentage displaced	19.0	8.4	10.6***	16.5	10.3	6.1**
		(2.2)	(1.4)	(2.6)	(2.4)	(1.4)	(2.7)
	Percentage household head	21.3	30.5	-9.2**	24.3	41.3	-17.0***
		(2.3)	(3.4)	(4)	(1.7)	(3.1)	(3.4)
	Percentage Single	5.6	11.7	-6.1***	10.7	12.0	-1.3
		(1.4)	(1.6)	(2.1)	(1.7)	(2.7)	(3.1)
	Percentage married of civil partnership	76.9	62.9	14.0***	77.2	67.4	9.8**
		(1.7)	(3.8)	(4.1)	(2.5)	(3.2)	(4)
Educational level (percentage)	None (level 0)	2.0	4.2	-2.1	2.3	2.7	-0.4
		(10)	(1.4)	(1.7)	(0.7)	(0.7)	(0.9)
	Primary incomplete (level 1)	18.1	18.6	-0.4	21.1	20.9	0.2
		(2.6)	(2.4)	(3.5)	(2.9)	(2.4)	(3.7)
	Primary complete (level 2)	13.5	13.8	-0.3	15.0	13.6	1.4
		(2.1)	(1.5)	(2.5)	(1.9)	(2.2)	(2.9)
	Secondary incomplete (level 3)	37.4	30.2	7.2*	35.3	30.7	4.6
		(2.5)	(3.3)	(4.1)	(2.6)	(2.2)	(3.4)
	Secondary complete (level 4)	24.0	25.4	-1.5	18.2	21.7	-3.5
		(2.9)	(2.6)	(3.8)	(2.1)	(2.5)	(3.2)
	More than secondary complete (level 5)	5.0	7.8	-2.8*	8.1	10.3	-2.2
		(1.2)	(1.2)	(1.7)	(2.6)	(1.5)	(2.9)
Dwelling characteristics	Household size	5.6	5.8	-0.2	5.6	5.8	-0.2
		(0.1)	(0.2)	(0.2)	(0.1)	(0.1)	(0.1)
	Number of people per room	3.2	2.9	0.3**	3.6	3.3	0.2*
		(0.1)	(0.1)	(0.1)	(0.1)	(0.1)	(0.1)
	Percentage with dirt floor	24.3	41.0	-16.7***	30.6	25.8	4.8
		(2.1)	(3.0)	(3.6)	(2.2)	(3.1)	(3.7)
	Percentage owning own house	82.7	58.7	24.1***	66.5	51.9	14.6***
		(2.0)	(3.0)	(3.5)	(2.2)	(2.7)	(3.4)
Public services (percentage)	Water by pipe	94.7	76.9	17.8***	91.9	83.4	8.5***
		(1.0)	(3.8)	(3.9)	(1.6)	(2.9)	(3.3)
	Sewer system	64.3	12.6	51.8***	67.3	25.5	41.8***
		(2.8)	(2.9)	(4.0)	(3.2)	(3.7)	(4.8)
	Does not have electricity	1.8	5.4	-3.6**	4.9	1.9	3.0*
		(0.7)	(1.5)	(1.6)	(1.4)	(0.7)	(1.6)
	Land phone	9.6	26.6	-17.0***	8.1	26.4	-18.3***
		(1.6)	(2.1)	(2.6)	(1.7)	(2.5)	(2.9)
Observations		342	334	676	346	368	714

Robust standard errors, clustered by session. * Significant at 10%; ** significant at 5%; *** significant at 1%.
For year 2008 we use Independent Cross section data.

In Table 3, we report the average level of cooperation in our sessions in 2007 and 2008, without distinguishing neighborhoods. The data from 2008 are divided between the subjects that had already played in 2007 and those that played the game for the first time. Several considerations are in order. First, although -as mentioned- the unique Nash equilibrium of that game assuming people maximize their own short-run earnings is that individuals invest their token in the private account, many individuals do deviate from the Nash equilibrium and contribute to the public good. Despite having a very low MPC and conducting the game in an

urban context, the overall level of cooperation we observe in our sample is similar to that observed in similar labs in the field. In the first round, only 14.8% and 12% of the sessions did not contribute at all. Second, there is a marked difference between the behavior of 'old and 'new' players.

Table 2. Demographic characteristics of the participants - 2007 and 2008 (Cont.)

		2007			2008		
	Variable	**Pozón**	**Ciénaga**	**Difference**	**Pozón**	**Ciénaga**	**Difference**
Assets (percentage)	Mobile phone	86.3	67.1	19.2***	76.3	68.7	7.6*
		(2.1)	(2.3)	(3.0)	(2.2)	(3.5)	(4.1)
	Bicycle	21.9	14.4	7.6**	19.4	15.2	4.1
		(3.3)	(1.8)	(3.7)	(2.5)	(2.4)	(3.3)
	Color TV	80.7	74.6	6.2*	84.4	84.5	-0.1
		(1.5)	(3.0)	(3.3)	(1.5)	(2)	(2.5)
	Washing machine	23.7	13.8	9.9***	27.5	27.7	-0.3
		(3.1)	(1.5)	(3.4)	(2.1)	(2.9)	(3.5)
	Sound player	39.8	25.1	14.6***	35.8	26.4	9.5***
		(2.8)	(2.8)	(3.9)	(2.2)	(2.6)	(3.3)
Income variables	Percentage unemployed	2.9	10.8	-7.9***	3.8	4.9	-1.1
		(0.9)	(2.3)	(2.5)	(1)	(1.1)	(1.5)
	Percentage with access to credit	67.8	66.8	1.1	72.3	69.3	3.0
		(2.4)	(2.5)	(3.4)	(1.8)	(2)	(2.7)
	Percentage with access to formal credit	21.1	22.2	-1.1	23.7	21.2	2.5
		(1.7)	(1.9)	(2.5)	(2.1)	(2.8)	(3.4)
	Percentage with food insecurity level (high)	9.6	9.9	-0.2	7.8	9.2	-1.4
		(1.4)	(1.9)	(2.3)	(1.6)	(1.6)	(2.2)
	Per capita monthly income (US$)	25.3	22.3	3.0***	30.3	27.1	3.2*
		(0.9)	(0.6)	(1.1)	(1.6)	(1.0)	(1.9)
Observations		342	334	676	346	368	714

Robust standard errors, clustered by session. * Significant at 10%; ** significant at 5%; *** significant at 1%. For year 2008 we use Independent Cross section data.

Virtually none of the 'old' players contributed to the public good and they almost uniformly played the Nash equilibrium. This result is likely to be driven by learning and by the fact that in the previous year contribution rates had been relatively low so that, even in *Pozón*, those who contributed to the public good received considerably less than the other players. This finding is also consistent with lab evidence on repeated VCM games in which, most of the times, in repeated rounds players tend to converge towards the Nash equilibrium.

Table 3. Behavior in the VCM - first round

	Baseline 2007	Follow-up 2008			
		Panel [a]	Difference [b]	Independent cross section	Difference [d]
Average percentage of contributors	20.0	7.3	-8.9**	29.3	9.3*
	(3.4)	(1.8)	(3.9)	(3.7)	(4.9)
Percentage of sessions with no contribution	14.8	28	9.1	10.5	-4.3
	(6.9)	(10.9)	(13.7)	(5.8)	(9.0)
Observations	676	439	878	714	1390

[a] Old participants in sessions with Old recruits, i.e. with 95% or more people were playing the game for the second time [b] Difference is calculated with respect to the same individuals' decisions in 2007. [c] Sessions with New recruits, i.e. with 95% or more people were playing the game for the first time. [d] Difference is calculated with respect to all decisions in baseline 2007. Robust standard errors in parenthesis, clustered by session. * Significant at 10%; ** significant at 5%; *** significant at 1%

In our analysis below, we present the results of the independent cross section only[57]. In Table 4, we report the descriptive statistics on several measures of 'social capital' that can be obtained from the post-game survey. These measures are divided into three groups: variables based on individual participation to civic association and neighborhood activities; variables reflecting voting behavior[58] and variables derived from answers to questions about trust and perception of cooperativeness[59]. The six columns of the table refer to the measures in Pozón, Ciénaga and the difference for the baseline and the follow-up.

In the first panel, we report the percentage of participants actively involved in neighborhood decisions or meetings on topics related to the community, and for active members in (at least one) civic organizations, the percentage of participants that attends to the meetings, are perceive her or himself as a leader, or as a decision maker and supports the association with money or voluntary work. In the second panel, we report the percentage of individuals who voted in local and presidential elections. In the third panel, we report statistics derived from a question on trust[60]. In addition, individuals may also have different perceptions on others' social preferences. They may perceive either a reciprocal (a dummy if most people help if others help), selfish (a dummy if most people care only about themselves), or cooperative (a dummy if most people help others unconditionally) behavior in their neighbors.

Compared to what was observed in other studies (see Latorre López, 2004 and Polania-Reyes, 2005) the participants reported higher levels of participation in organized groups. However, our measures of civic participation are significantly lower in the follow up. Voting behavior in 2008 is significantly higher than in 2007 for both local and presidential elections, which is consistent with some studies on voting behavior and CCT (see Zárate et al., 2013 and Nupia, 2012; see footnote 37 for evidence on other countries). In the first two panels, the time trend in social capital measures is the same in both years. In the local context, voting participation is different in 2007 although we are able to argue that this measure is dependent on the time of the survey: in 2007, three months passed after the local elections and it was more than a year we ask the same question in 2008. Direct measures of trust are not significantly different across time and yield particularly a low estimate of trust. However, the sign of the difference is consistent with what we would expect: the program might have reduced the percentage of people who think no one can be trusted and the percentage of people who think that most of the people in the community are selfish.

In Table 5 we present the average number of friends, acquaintances and connections (the sum of relatives, friends and acquaintances) each participant reports in the session where she is

[57] In order to avoid confounding the impact of the program with learning effects. Subjects learn from other's behavior, and update their beliefs about what other will do the next time they have the chance to play. This effect is particularly strong when the subjects know each other before entering the lab (a key issue of experiments in the field). In 2007, it is very likely that subjects met and discuss the results of the activity once they left the session. Participants in 2007 that participated again in 2008 have a different characteristic from those who face the game for the first time: they may contribute as the experimental evidence affirms that contributions decrease over time but increase in the presence of communication, they may also contribute if they contributed before.

[58] The last local elections in Cartagena (Governor, mayor, members of the *Asambleas Departamentales*, Municipal council and *Juntas Administradoras Locales*) were held three months after the inscription in the program (Oct. 2007) and eight months before the beginning of the program in *Pozón* (April 2003). Presidential elections were held sixteen months after the beginning of the program in Pozón, in May 2006.

[59] There is a risk that the attitudinal measures of social capital were influenced by the experimental games (Carpenter, 2002) as the players answered the questionnaire after they knew the outcome of the games and they possibly might have expressed their emotions in their answers on perception of trust and cooperation.

[60] This question was adapted from the WVS source. We added an alternative "few people can be trusted".

playing. We also report a feature of the network in the session, the friendship, acquaintanceship and connectivity densities, measured as the ratio of the total number of identified specific links in the session and the total possible number of specific links among connected people, i.e. those individuals that are identified as an acquaintance at least once by other player. In addition, we present a measure of leadership as the percentage of players identified as leader in each session, at least by one different player in the session.

Table 4. Traditional social capital measures - independent cross section

	Percentage		2007	2008	Difference	Standard errors
Civic Participation	Participation in neighborhood decisions		54.3	40.9	-13.4***	3.8
	Participation in the neighborhood meetings		61.8	47.9	-13.9***	3.0
	Membership in at least one organization		38.6	26.8	-11.9***	3.7
	If is member of any civic association	Attendance to the meetings	38.5	26.3	-12.1***	3.7
		Decision maker	33.6	23.5	-10.1***	3.5
		Leader	18.5	9.8	-8.7***	2.1
		Supports with money or voluntary work	35.7	23	-12.7***	3.4
Voting behavior	Voted in local elections (2003 and 2007)		52.4	72.4	20.0***	3.0
	Voted in presidential elections (2006)		67	77.7	10.7**	4.4
Trust and cooperation perception	Trust	Most people	6.7	6.6	-0.1	1.6
		Few people	62	62	0.0	2.9
		None	31.3	31.4	0.1	2.8
	Perception within the community	Cooperation	26.5	32.3	5.8***	2.2
		Reciprocity	24.2	14	-10.2***	2.3
		Selfishness	49.4	53.7	4.3	2.8
Observations			676	714	1390	

Robust standard errors of the difference, clustered by session. * Significant at 10%; ** significant at 5%; *** significant at 1%.

Participants in Pozón in 2007 attended sessions with higher levels of connectivity but these differences are gone in the follow up. We are not able to explain rigorously these differences as we may find that the program balanced the difference between networks in the sessions in the follow up. We may also say that the recruitment process in 2007 facilitated participants to attend in groups of people with pre-existing links. For example, in Ciénaga in 2007, participants had fewer friends than in Pozón 2007 and 2008 but in 2008 participants in Ciénaga attended the sessions with more friends.

Table 5. Networks characteristics across sessions - independent cross section

	2007	2008	Difference	Standard errors
Average number of friends	1.0	1.5	0.4*	0.2
Average number of acquaintances	7.7	0.4	-7.3***	0.6
Average number of connections	8.9	2.0	-6.9***	0.7
Friendship density[a]	0.1	0.1	0.1***	0.0
Acquaintanceship density	0.4	0.0	-0.3***	0.0
Connections density	0.4	0.2	-0.3***	0.0
Percentage of players identified as leader	33.4	18.1	-15.4***	3.3
Observations	676	714	1390	

[a] Network density is calculated as the number of connections /(number of connected people)*(number connected people -1). For example, Friends density is calculated as the number of friends/(number of people with friends)*(number of people with friends -1). Robust standard errors of the difference clustered by session. * Significant at 10%; ** significant at 5%; *** significant at 1%.

We also find that the percentages of players identified as leaders in the session is significantly lower in the follow up, suggesting that a certain fragmentation affects the patterns of leadership consolidation in the social context of the study. Interestingly, the rate of reported

leaders in baseline and follow up is significantly higher than the proportion of ML (participants who declared to have been elected FA beneficiary representatives) in the cross-section sample (9.0% and 5.1% respectively)[61].

Surveys record stated preferences while experiment outcomes provide revealed preferences. The experimental literature on economic experiments and surveys does not find a consistent answer on whether these two are positively correlated[62]. On one hand Gächter et al. (2004) and Capra et al. (2008) find a positive correlation between stated and revealed preferences in the public goods game and trust game. On the other hand Cardenas et al. (2013) find no evidence of correlation between participation in a charity and contribution in a public goods game[63].

4. Identification issues and empirical specification

To establish the relationship between the expansion of the CCT program and social capital we compare the behavior in the game of individuals who had been exposed to the program to those who had not. In July 2007, the program had been operating in the *Pozón* for over a year, while it had not been operating in *Ciénaga*. It should be stressed that the expansion to *Ciénaga* had already been decided so that the FA office could provide us with lists of potential beneficiaries in *Ciénaga*. In July 2008, we conducted the same game in both neighborhoods, when the program was operating in both. In addition we gathered information on individual and household level variables, as well as on the links (relatives and friends) among the participants to the sessions. By the time we conducted the follow up experiments, *Ciénaga* had been already targeted for over 12 months.

The expansion to *Ciénaga* allows us to implement a difference in differences (DiD) strategy to control for the possibility that the two neighborhoods differ in some unobserved dimension, provided that such 'unobservable' differences do not change over time. Although this is an untestable assumption we show below that most structural individual and household characteristics were stable within each neighborhood during the evaluation time.

The first identification issue raised by our study is the lack of randomization in the allocation of the program across neighborhoods. Based on conversation with program Administrators, it was felt that the initial allocation of FA was not driven by specific pre-existing differences between neighborhoods.

Although the availability of information on a wide range of pre-intervention variables allows us to control for the possibility that the two neighborhoods differed in the level of social capital for reasons different from the operation of the program, it is possible that there are unobserved differences that we could not control for. This motivated us to perform a follow up study a year after the first stage of the study. As discussed in the introduction, in 2008

[61] We find that 84.2% (for baseline) and 46.2% (for follow up) among those identified as leaders in the session are MLs.

[62] This could be explained with what Carpenter (2002) refers to as the idealized persona bias, i.e. a desire to appear as the person one would like to be rather than what they are, and/or the surveyor effect, i.e. the desire to answer in accordance to what one thinks the surveyor would like to hear or see.

[63] Our study is not exceptional, we present a comparison between behavior in the game and the traditional measures of social capital with traditional measures exhibit a weak correlation with such behavior. In fact, the correlations we find are not stable over time, which reinforces the concern of using these measures in the analysis of social capital.

both neighborhoods were exposed to the program so that, under some assumptions, it is possible to use the data from our second experiment to estimate these unobserved differences and correct the estimates derived from considering the difference between treatment and control in 2007.

The approach we take is essentially a DiD one. The only difference relative to a standard situation where at baseline both treatment and control are not exposed to the program while at follow up only the treatment are, is that in our case only one neighborhood was exposed at baseline, while both were exposed to the treatment at follow up. Unlike in the standard case where unobserved differences between the two groups are estimated using baseline data, we use the follow up data to do the same thing.

For our approach to be valid we need a number of assumptions. First, we need to assume that, conditional on observed variables, unobserved differences between the two neighborhoods are constant over time and enter additively. This is what is usually referred to as the common trend hypothesis. Unfortunately, it is not possible to test this assumption because of the lack of retrospective information at the individual or neighborhood level. However, we discuss below why we think this is a reasonable assumption to make on the basis of the available descriptive statistics.

Second, we need to assume that, after a few months of exposure, the impact of the intervention on social capital does not change with exposure. That is, if the program affects social capital, it does so by moving, after a few months, the social capital to a new level. This assumption is necessary to allow us to identify the unobserved differences in social capital between the two neighborhoods using the 2008 data when the program had operated for over 2 years in Pozón and only for just over 11 months in Ciénaga. Such a hypothesis is plausible as intensive interaction among beneficiaries starts early in the process. We assume that the interactions between beneficiaries in the first year of exposure to the program are sufficient to bring the level of social capital at a new equilibrium level that does not change after this first initial period.

To improve efficiency of our analysis, we can also control for a number of variables that are unaffected by the program. The common trend assumption we make to use the DiD approach, has to hold conditional on these variables.

In order to model the probability to contribute to the public good, we run a simple probit regression, where a single value index Y_{ist} is modelled as follows:

$$(1) \qquad Y_{ist} = \alpha + \beta X_{ist} + \gamma D_{i07} + \delta T_{i07} + \lambda(1 - D_{i07} * T_{i07}) + \theta S_{st} + v_{st} + \varepsilon_{ist}$$

where Y_{ist} is the individual decision to contribute to the group account in session s at time t; D_{i07} is a dummy for Ciénaga (which is the *control* group in 2007); T_{i07} is a dummy for the year 2007, when Ciénaga was not exposed to the program; X_{ist} are individual observable characteristics, S_{st} are session level characteristics and ε_{ist} are i.i.d. Gaussian distributed error term with mean zero and variance $\sigma_{\varepsilon}^2 = 1$, independently of v_{st}, which are iid, $N(0, \sigma_v^2)$. The parameter λ in the expression above captures the effect of the program on the single value index Y_{ist} and determines the impact of the program on the probability to contribute to the public good. The time dummy captures common trend effects by exploiting the difference in contribution rates between 2007 and 2008 in *Pozón, (*assuming that any FA impact had reached a stable level already in 2007); the neighborhood dummy captures fixed

effects by exploiting the difference in contribution rates between *Pozón* and Ciénaga in 2008 (assuming that any FA impact had reached a stable level in both neighborhoods then).

Another potentially important issue is related to our sampling method. As we discussed above, due to logistical and data availability constraints, the recruitment approach was different across in 2007 and 2008 and, in 2007 was different between treatment and control neighborhoods. In both neighborhoods and both data collection stages the sample of participants in session includes individuals that self-selected for participation into the experiment by responding to an invitation. Hence the study is not based on a pure random sample and results might be non- representative of the whole beneficiary population. However the design was established to ensure that the self-selection bias affected the treatment and control group in a similar way at each stage of the study. Table 2 above constitutes an important and reassuring piece of evidence in this respect. As we mentioned, whilst the two neighborhoods appear to be different in a number of observable dimensions, these dimensions are virtually identical over time. If the selection process had been substantially different in Ciénaga in 2007 and 2008 and had been the same in Pozón, then this would have resulted in the observed differences between the two samples to change over time.

We conclude this discussion with a consideration about statistical inferences. Observations within a small neighborhood could be correlated because neighbors share common characteristics besides the assignment into treatment and control. However, the DiD approach does take into account the presence of a systematic difference between the two subsamples. As it turns out, for many outcomes, the intra-class correlation coefficient within neighborhoods is relatively low at 0.010[64]. On the other hand, the behavior of participants at a given session is affected by many common factors, ranging from the identity of the game master, to the time of day to the general atmosphere in the session. Therefore we account for associations among observations within sessions (clusters) and cluster our standard errors at the session level.

5. Do CCTs build social capital?

We identify the effects of FA on our experimental measure of social capital by comparing first the contribution rates to the group account in our VCM in the two neighborhoods in 2007 and 2008. Then, we analyze the impact of the program on different and more traditional measures of social capital (Chattopadhyay and Duflo, 2004, Jones and Olken, 2005, Beaman et al., 2009, Banerjee et al., 2013).

i. Identifying the effect of FA on the willingness to contribute to a public good

In Table 6, we pool together the 2007 and 2008 data to identify the effect of FA on the probability to contribute to the group account, after controlling for a variety of observable variables (the complete set of results is presented in Table 8), using the probit model in equation (1) in all four columns. In Column I we do not add any controls. In column II we control only for individual characteristics such as the age and education level of the participant. In Column III we control for session characteristics. An issue that worried us

[64] Individuals within neighborhoods are no more similar than individuals in different neighborhoods (in part for our sample size), and we effectively assigned 1390 individuals to treatment or control. We can reject that we have only two independent observations.

considerably on the field was the possibility that individuals who played in early sessions would 'contaminate' other individuals that were about to enter subsequent session by talking to them and commenting on the game. While we tried to avoid these contacts as much as possible and instructed the subjects not to talk to subsequent players, some contacts were unavoidable. For this reason, we control for the sequence order in which a particular session is played in a day and in the full sequence of sessions and also for results in previous sessions[65]. We do find these session effects to be significant. However, their introduction does not change the size of the marginal impact, partly because these effects were present in both treatment and control sessions.

Under the assumptions discussed above, the coefficient λ in equation (1) can be interpreted as the impact of the program and its estimates are reported in the third row of the table. The impact is large and significant at the 1% level.

The positive coefficient on the 2007 dummy indicates (under the maintained assumptions) the presence of a negative common trend in our measure of social capital. Finally, the positive coefficient on the Ciénaga dummy indicates that unobserved differences in social capital (not related to the presence of the program) are higher in Ciénaga than in Pozón.

The coefficients in the first column are the results of mean participation rates that are higher in Pozón than in Ciénaga (when the result was operating in both neighborhoods) and they flip in 2008 when the program was operating in both. Therefore the bias in estimating the impact of the program using only the data from 2007, as in Attanasio et al. (2009) would be negative. Notice, however, that the coefficient on the Ciénaga dummy becomes statistically insignificant once we control for observable variables. Notice also that if one thinks that the impact of the program takes more than 12 months to develop and that in 2008 its impact was stronger in Pozón than Ciénaga because the program had operated longer in the former than in the latter, then the estimates we present in Table 6 would be biased downwards.

Table 6. Program impact on cooperation - first round

Dependent variable: 1 if player contributed to the group account in the first round				
Independent Variables	I	II	III	IV
Dummy *Ciénaga*	0.105*	0.084	0.048	0.064
	(0.060)	(0.057)	(0.062)	(0.052)
Dummy 2007	0.089*	0.090*	0.064*	0.096***
	(0.050)	(0.049)	(0.037)	(0.033)
Program impact	0.306***	0.293***	0.241***	0.233***
	(0.044)	(0.046)	(0.043)	(0.034)
Basic controls - participant characteristics	No	Yes	Yes	Yes
Session characteristics	No	No	Yes	Yes
Session composition (info within session)	No	No	No	Yes
Observations	1390	1384	1384	1384

Marginal effects reported. Robust standard errors, clustered by session, in brackets after probit estimation. * significant at 10%; ** significant at 5%; *** significant at 1%

In the specifications in Column III and IV, the point estimates of the impact of the program are now lower, but they are still significantly different from zero at the 5% level. The main message

[65] Particularly, in order to control for contiguous sessions correlation, we include in specification (IV): a) a dummy for the first session each day; b) a variable capturing the deviation from the neighborhood mean of the average contribution to the public account in the previous 2 sessions, and c) the order (number) in which the session came about in the full sequence of sessions within the neighborhood.

of the table, therefore, is that FA had a strong impact on social capital; at least as measured by the behavior in the game we played in Cartagena. This result seems to be robust to the presence of systematic differences between the two neighborhoods we have compared, both in terms of observable and unobservable variables. Given the magnitude and robustness of the estimate, it doesn't seem plausible that this result is driven exclusively by differential trends in unobservable across Ciénaga and Pozón.

ii. Impact on traditional measures

We now analyze the impact of the program on the measures of social capital we summarized in Table 4 and Table 5. We implement equation (1) to model the probability that an individual engages in a 'pro-social' activity (e.g. participation in neighborhood decisions) with a probit regression[66] in which we include the same individual control variables that we considered in the range of specifications for the decision to contribute in the game in Table 6.

Table 7 reports the impact of the program (the coefficient λ) on each of these measures of social capital. A few social capital measures are statistically significant across both specifications. The clearest effect is on 'individual social integration' variables as well as on 'voting behavior'. The effect on 'civic participation' is still positive but mixed: it is still positive except for participation in at least one civic organization (where coefficients are negative but not significant', consistent with Pellerano (2008) who, considering the rural program, observed that participation into the EC seemed to substitute for participation in other organizations). Finally there is a clear effect on perceptions about trust and cooperation. Whilst the program does not seem to create positive perceptions about other people's attitudes towards trust and cooperation, it does seem to reduce the negative ones. This is a noteworthy result that sheds light on the manner in which the CCT can affect beliefs.

Overall, Table 7 depicts a picture that complements the one arising from Table 6 in terms of impact of the CCT program[67].

6. The determinants of social capital

The evidence we have presented so far seems to indicate that FA had a strong impact on different dimensions of social capital. In the process of getting this result, in section 5, we have controlled for several variables, at the individual, household and session level, to ensure that the results we obtained could not be explained by differences in these variables between the two neighborhoods considered. The coefficients on these variables are interesting on their own, as they are informative about the main determinants of cooperative behavior. In this section, we discuss the coefficients we obtained on the various variables we considered as determinants of the behavior in the game.

[66] For the measures on Individual social integration we use a linear regression model.

[67] The question arises as to whether Table 7 says the effect of the CCT is indirect, via the variables captured by the survey, or direct. From Table A2 we observe there is no significant link between the survey variables and the behavioral one, which provides some external validity to the claim that the program has an impact on cooperation. Having discussed the plausibility for lack of correlation between stated and revealed preferences, this should not raise a concern.

Table 7. Program impact on traditional measures of social capital

Probit marginal effect of the program impact [a]			I	II
Civic Participation	Membership in at least one organization		-0.070	-0.116*
			(0.071)	(0.069)
	Participation in neighborhood decisions		0.236***	0.221***
			(0.049)	(0.054)
	Participation in the neighborhood meetings		0.158***	0.111**
			(0.046)	(0.050)
Voting behavior	Vote Local Elections (2003 and 2007)		0.159***	0.134**
			(0.055)	(0.058)
	Vote Presidential Elections (2006)		0.171***	0.095
			(0.061)	(0.062)
Trust and cooperation perception	Trust	Most people	0.005	0.003
			(0.029)	(0.027)
		Few people	0.127**	0.116*
			(0.056)	(0.062)
		None	-0.136***	-0.122**
			(0.051)	(0.057)
	Perception within the community	Cooperation	0.082**	0.099**
			(0.040)	(0.043)
		Reciprocity	0.013	0.009
			(0.047)	(0.045)
		Selfishness	-0.104**	-0.118**
			(0.050)	(0.052)
Individual Social Integration	Number of player's relatives		0.034	-0.185***
			(0.023)	(0.049)
	Number of player's friends		1.255***	0.671*
			(0.139)	(0.380)
	Number of player's acquaintances		1.097***	2.350**
			(0.347)	(1.026)
	Number of player's connections		2.386***	2.836***
			(0.352)	(1.028)
Player is identified as leader in the session			0.136***	0.122***
			(0.041)	(0.044)
Basic Controls - Participant's characteristics			No	Yes
Session Characteristics			No	No
Session composition (info within session)			No	No
Observations			1390	1384

Coefficients for the marginal effects in probit specifications I-II, the dependent variable being the one listed on the left hand side. Robust standard errors, clustered by session, in brackets after probit estimation. *Significant at 10%; **significant at 5%; *** significant at 1%. [a] Coefficients for the measures of individual social integration are OLS estimates

i. Individual and session variables effects on social capital

In the first panel of Table 8, we report the coefficients on the individual and household level variables in the regressions that were used to compute the impact of the program as reported in Table 6. Age is an important determinant of cooperation: older players seem to be more likely to contribute to the group account. The few men who participated (mainly in 2007) in the game were less likely to cooperate. As for household level variables, we find that players with a large household size tend to cooperate more. Interestingly, we do not find any education effects and very limited effects of other household level variables, including various wealth indicators.

Table 8. Controls for willingness to cooperate in the first round (Table 6) - independent cross section

Dependent variable: 1 if player contributed to the group account		II	III	IV
	1 if the player is a woman	0.118**	0.156***	0.158***
		(0.058)	(0.030)	(0.028)
	If the player has a partner	-0.002	0.006	0.008
		(0.036)	(0.033)	(0.033)
	Level of education (0 to 5)	-0.001	-0.007	-0.005
		(0.010)	(0.009)	(0.009)
	Household size	0.016***	0.012**	0.013**
		(0.006)	(0.006)	(0.006)
	Ground floor (house)	-0.045*	-0.031	-0.030
		(0.025)	(0.023)	(0.023)
	Number of years living in the neighborhood	0.001	0.001	0.001
		(0.001)	(0.001)	(0.001)
	1 if the player's home has water pipe access	0.043	0.038	0.044
		(0.035)	(0.032)	(0.031)
	1 if the player's home has sewage	-0.039	-0.031	-0.024
		(0.029)	(0.028)	(0.027)
	1 if the player's home has no electricity	0.128*	0.086	0.096
		(0.072)	(0.071)	(0.072)
Basic controls - participant characteristics	1 if the player is the head of household	0.019	0.022	0.024
		(0.039)	(0.039)	(0.040)
	Age	0.005***	0.005***	0.005***
		(0.001)	(0.001)	(0.001)
	1 if the player is unemployed	-0.017	0.007	0.008
		(0.057)	(0.058)	(0.058)
	1 if the player has her own housing	-0.041	-0.049*	-0.048*
		(0.027)	(0.025)	(0.025)
	1 if player belongs to the fifth quintile per capita income	0.047	0.036	0.041
		(0.031)	(0.030)	(0.030)
	1 if the player has a phone	-0.001	-0.001	-0.005
		(0.038)	(0.038)	(0.038)
	Number of rooms in the house	-0.018	-0.011	-0.010
		(0.014)	(0.014)	(0.015)
	1 if the player is displaced (self-declared)	0.025	0.015	0.013
		(0.032)	(0.032)	(0.032)
	1 if the player declares she understood everything	-0.048	-0.043	-0.044
		(0.029)	(0.030)	(0.030)
	1 if the household receives support from another institution	-0.063**	-0.054**	-0.052*
		(0.029)	(0.027)	(0.027)
Observations		1384	1384	1384

Marginal Probit for II-IV. Marginal effects reported. Robust standard errors, clustered by session. * Significant at 10%; ** significant at 5%; *** significant at 1%. [a] Calculated as the Mean of previous 2 Sessions' Deviation from the average contribution in the neighborhood.

Several of the session level variables are significant determinants of cooperation. Whilst men are less likely to contribute, the presence of a man in the session seems to induce cooperation. Players in the first session of the day were much more likely to cooperate, while the total number of players in a session had a negative effect on cooperation. Finally, when we consider (in Column IV) another set of variables controlling for the *composition* of the sessions, we find that participants in a session with a high percentage of the highest level of education (i.e. secondary education) within the session and a session socially distant with respect to other sessions in the neighborhood (in terms of the level of education of its

participants) will be more willing to contribute to the public good[68].

Table 8. Controls for willingness to cooperate in the first round (Table 6) - independent cross section (Cont.)

Dependent variable: 1 if player contributed to the group account		II	III	IV
Session Characteristics	Number of players in session		-0.030***	-0.037***
			(0.009)	(0.008)
	1 if there is at least one man in the session		0.166***	0.201***
			(0.043)	(0.044)
	Experimenter n°2 (female) in 2007		-0.049	-0.071*
			(0.052)	(0.043)
	Experimenter n°3 (male) in 2007		0.121**	0.135***
			(0.047)	(0.043)
	Experimenter n°2 (female) in 2008		0.099*	0.096**
			(0.058)	(0.049)
	First session in the day		0.080**	0.095***
			(0.040)	(0.035)
	Behavior in the 1st round of the last two sessions [a]		-0.316*	-0.226*
			(0.162)	(0.124)
	Number of sessions held before (t)		-0.020***	-0.019***
			(0.003)	(0.003)
Session composition	Percentage of participants with less than secondary complete			0.370***
				(0.109)
	Mean absolute deviation - level of education in the neighborhood			0.198*
				(0.103)
Observations		1384	1384	1384

Marginal Probit for II-IV. Marginal effects reported. Robust standard errors, clustered by session. * Significant at 10%; ** significant at 5%; *** significant at 1%. [a] Calculated as the Mean of previous 2 Sessions' Deviation from the average contribution in the neighborhood.

We also note that, although the three game administrators (who were rotated across the sessions in both neighborhoods) were instructed to read exactly the same protocol and their behavior in the sessions differed (necessarily) only when answering the questions posed by the participants, we find strong 'administrator' effects.

ii. Leadership and social capital

An important feature of the FA program is the importance that the role of leaders acquires in the community activities and therefore their role in strengthening social ties with the communities and becoming channels of wellbeing improvement (Latham and Saari, 1979, Bass, 1991)[69]. In this subsection, we investigate whether the presence of a ML in the session affected individual decisions. Table 8 presents the results we obtain if we consider, among the determinants of individual behavior, the presence of a ML among the participants. Cooperation is positively associated with the presence of at least one ML in the session. The marginal effect is large at around 13.2%, when we control for session characteristics and

[68] Despite considering our network measures as measures of social capital instead of variables that explain cooperation, we explored the role of preexisting family and friend networks among the beneficiaries in the decision to contribute. Among all the specifications we attempt by adding different combinations of the network measures to specification (IV) in Table 6, only found a significant association between Friendship density of the session and the decision to contribute (i.e. marginal effect of 0.538, significant at 1%).

[69] Brune and Bossert (2009) find a positive effect of an intervention based on leadership development on a wide array of social capital measures from the World Bank's Social Capital Assessment Tool in Nicaragua. Macours and Vakis (2014) found that social interactions with nearby leaders positively affected human capital and productive investments in a CCT in Nicaragua.

composition (specifications VII and VIII), and above 16% when we do not control for session these variables (see specifications V and VI). The Program impact has a lower coefficient but still positive and significant. We can propose that one of the channels through which cooperation succeeds in the community is their leader. The results we obtain on the other control variables are not different from those presented in Table 8 are and are reported, for completeness, in Appendix C in Table C2.

In addition, being an ML may be a determinant of the attitudes to cooperate, not only to the beneficiary that was elected as a ML but also to those who may be considered leaders by the group despite of not having an official title. In Table C3 in the appendix we present the results when we include whether the individual is considered as a Leader in the group and the percentage of participants who consider this particular individual as one of their leaders in the group.

Table 9. Role of the leader in the first round

Dependent variable: 1 if player contributed to the group account in the first round				
Independent Variables	V	VI	VII	VIII
Dummy *Ciénaga*	0.103*	0.083	-0.003	0.028
	(0.056)	(0.051)	(0.042)	(0.039)
Dummy 2007	0.056	0.059	0.060*	0.091***
	(0.050)	(0.048)	(0.036)	(0.031)
Program impact	-0.202***	-0.188***	-0.107*	-0.121**
	(0.070)	(0.070)	(0.065)	(0.056)
1 if there is at least a ML in the session	0.165***	0.166***	0.144***	0.132***
	(0.062)	(0.058)	(0.040)	(0.041)
Basic controls - participant characteristics	No	Yes	Yes	Yes
Session characteristics	No	No	Yes	Yes
Session composition (information within session)	No	No	No	Yes
Observations	1,390	1,384	1,384	1,384

Marginal probit. Marginal effects reported. Robust standard errors, clustered by session, in brackets after probit estimation. * Significant at 10%; ** significant at 5%; *** significant at 1%. We don't use the dummy indicating whether participant is a ML: it is collinear with program impact and the coordinator 2 in 2007. We would have to exclude the data from *Ciénaga* in 2007 as the program was not operating there and, as a consequence, there were no ML in those sessions and the DID would not be feasible.

7. Conclusions

Our main purpose is to examine, by combining survey and experimental methods, the effect of a social intervention on individual and group behavior and their ability to obtain better social and economic outcomes. There are two main difficulties in testing rigorously the hypothesis that CCTs increase social capital. The first lies in the difficulty in obtaining a quantitative measure of social capital. The second is the standard evaluation problem of observing the counterfactual to a given intervention.

We addressed the first difficulty implementing a public goods game. Our main measure of social capital is based on the behavior during this game, whose rewards, for the individuals in our sample, are salient. Attanasio et al. (2009) started our analysis by comparing the level of cooperation in two similar neighborhoods in 2007 when the CCT program we have been studying was operating only in one of them and show that the level of cooperation we observe in the 'treatment' community is considerably higher than in the 'control' community. In 2008, after the program had been expanded to the control community, we retake the measurements taken in 2007 and find no strong difference between cooperation levels in the two neighborhoods. By using a difference in difference approach, we control for unobserved

differences in the social fabric of the two communities and find evidence to support the hypothesis that the introduction of the CCT we have been studying did indeed increase social capital.

The other focus of the paper is on how to measure social capital. While we implemented a VCM game and used observed individual behavior in the game as our main measure, we also collected information on self-reported behavior along a number of dimensions that have been used in the literature as measures of social capital. We show that while these measures roughly correlate with ours, they are also affected by measurement error, which makes the identification of the impacts of CCTs on social capital harder to detect. Despite this, with our DiD approach we found a positive effect of the CCT on participation in neighborhood decisions, participation in the neighborhood meetings and a negative effect of the CCT on the perception that "it is not possible to trust in anybody" and in the perception that "in the community people only care about themselves". We only found a weak effect of the CCT on the number of connections subjects have within the session. We also explore the role of preexisting family and friend networks among the beneficiaries in an alternative effect of a CCT. Although it has been documented how pre-existent networks may determine the effect of a CCT program on the main outcome (Angelucci and De Giorgi, 2009, Angelucci, et al., 2010, Macours and Vakis, 2014), it has not been the case with the effect of our particular CCT.

It must be noted that the external validity of our results might be limited by some of the sample selection issues we discussed above: our sample was made of self-selected individuals that responded to an invitation to participate to game sessions and the formation of groups was also left to individual self-selection into sessions depending on the timing of arrival at the game site. Despite these caveats, and whilst results are not representative to the whole population of beneficiaries of FA, they provide compelling evidence about the potential influence of CCTs on social capital formation and the mechanisms through which such effect is likely to realize.

The effect that CCT programs such as FA might have on social capital, probably because of the many social activities linked to the program, has long been debated. Some policy makers even hypothesized that the program might have a negative impact on social capital creating division between beneficiaries and (marginal) non-beneficiaries. Our evidence represents one of the first quantitative assessments of the impact of a CCT on social capital and provides strong support for the hypothesis that social capital is improved. The policy implications of these findings are obvious. This impact should be added to the benefits that Conditional Cash Transfers could deliver. Of course, one has to be careful and understand that the manner in which a CCT is delivered might be key to the social capital impacts we see.

8. References

Adato, Michelle. 2000. "Final report: The impact of Progresa on community social relationships", International Food Policy Research Institute, Washington, D.C. http://evaluacion.oportunidades.gob.mx:8010/es/wersd53465sdg1/docs/2000/ifpri_2000_community_level_impacts.pdf

Adato, Michelle. Integrating survey and ethnographic methods to evaluate conditional cash transfer programs. Vol. 810. Intl Food Policy Res Inst, 2008.

Adato, Michelle, John Hoddinott, and Lawrence James Haddad. *Power, politics, and performance: community participation in South African public works programs*. Vol. 143. Intl Food Policy Res Inst, 2005.

Adato, Michelle, and John Hoddinott. 2010. Conditional cash transfers in Latin America. International Food Policy Research Institute. http://www.ifpri.org/sites/default/files/publications/conditional_cash_transfers_in_latin_america.pdf

Alzua, Maria Laura, Juan Camilo Cardenas, and Habiba Djebbari. 2014, Community mobilization around social dilemmas: evidence from lab experiments in rural Mali. No. 0160. CEDLAS, Universidad Nacional de La Plata.

Andreoni, James. 1995. "Cooperation in Public Goods Experiments: Kindness or Confusion." American Economic Review, 85(4), pp. 891-904.

____. 1988. "Why Free Ride? Strategies and Learning in Public Good Experiments." Journal of Public Economics, 37, pp. 291-304.

Angelucci, Manuela, and Giacomo De Giorgi. "Indirect effects of an aid program: How do cash transfers affect ineligibles' consumption?." The American Economic Review (2009): 486-508.

Angelucci, Manuela, Giacomo De Giorgi, Marcos A. Rangel, and Imran Rasul. 2010 "Family networks and school enrolment: Evidence from a randomized social experiment." Journal of Public Economics 94.3: 197-221.

Arriagada I, Mathivet C. 2007. Los programas de alivio a la pobreza Puente y Oportunidades. Una mirada desde los actores. *Polít. Soc. Ser.* 134. Santiago de Chile: CEPAL. 40 pp.

Attanasio, Orazio; E. Battistin; E. Fitzsimon; A. and Mesnard and M Vera-Hernandez. 2005. "How Effective Are Conditional Cash Transfers? Evidence from Colombia," In IFS Briefing note

Attanasio, Orazio, Emla Fitzsimons, Ana Gomez, Martha Isabel Gutiérrez, Costas Meghir and Alice Mesnard, 2010. Children's Schooling and Work in the Presence of a Conditional Cash Transfer Program in Rural Colombia. Economic Development and Cultural Change, Vol. 58, No. 2 (January), pp. 181-210

Attanasio, Orazio, Erich Battistin and Alice Mesnard, 2012a. "Food and Cash Transfers: Evidence from Colombia," Economic Journal, Royal Economic Society, vol. 122(559), pages 92-124, 03.

Attanasio Orazio and Pellerano Luca. 2012b. "Impactos de largo plazo del programa familias en acción en municipios de menos de 100 mil habitantes en los aspectos claves del desarrollo del capital humano" Final Report. DNP. Bogota. https://sinergia.dnp.gov.co/Sinergia/Archivos/92646308-df90-4107-a074-8bf62ea9e862/Eval_Familias%20Acci%C3%B3n%20Largo%20plazo.pdf

Attanasio, Orazio; Abigail Barr; Juan-Camilo Cardenas; Garance Genicot and Costas Meghir. 2012b. "Risk Pooling, Risk Preferences, and Social Networks." American Economic Journal: Applied Economics, 4, no. 2: 134-167.

Attanasio, Orazio; Luca Pellerano and Sandra Polanía-Reyes. 2009. "Building Trust? Conditional Cash Transfer Programs and Social Capital." Fiscal Studies, 30(2), pp. 139-77.

Avdeenko, Alexandra, and Michael J. Gilligan. "International interventions to build social capital: evidence from a field experiment in Sudan." (2014).

Baez, Javier Eduardo, and Adriana Camacho. "Assessing the long-term effects of conditional cash transfers on human capital: Evidence from Colombia." World Bank Policy Research Working Paper Series, Vol (2011).

Barr, Abigail, Frederick Mugisha, Pieter Serneels, and Andrew Zeitlin, 2012. "Information and collective action in the community monitoring of schools: Field and lab

experimental evidence from Uganda," Unpublished. Retreived at: https://dl.dropboxusercontent.com/u/11087574/papers/m%26m.pdf
Barr, Abigail, Truman Packard, and Danila Serra. 2014. "Participatory accountability and collective action: Experimental evidence from Albania." European Economic Review 68: 250-269.
Bass, Bernard M. "From transactional to transformational leadership: Learning to share the vision." Organizational dynamics 18, no. 3 (1991): 19-31.
Bowles, Samuel and Herbert Gintis. 2004. "The Evolution of Strong Reciprocity: Cooperation in a Heterogeneous Population." Theoretical Population Biology, 65, pp. 17-28.
Brune, Nancy E., and Thomas Bossert. "Building social capital in post-conflict communities: Evidence from Nicaragua." Social Science & Medicine 68, no. 5 (2009): 885-893.
Camacho, Luis A. 2014. The Effects of Conditional Cash Transfers on Social Engagement and Trust in Institutions: Evidence from Peru's Juntos Program (October 31). Discussion Paper 24/2014, German Development Institute, Bonn. Available at SSRN: http://ssrn.com/abstract=2512666 or http://dx.doi.org/10.2139/ssrn.2512666
Camacho, Adriana, and Daniel Mejía. (2013) Las externalidades de los programas de transferencias condicionadas sobre el crimen: El caso de Familias en Acción en Bogotá. Inter-American Development Bank. Working paper IDB-WP-406.
Camerer, Colin and Ernst Fehr. 2004. "Measuring Social Norms and Preferences Using Experimental Games: A Guide for Social Scientists," In Foundations of Human Sociality: Economic Experiments and Ethnographic Evidence from Fifteen Small-Scale Societies, ed. J. Henrich, S. Bowles, R. Boyd, C. Camerer, E. Fehr and H. Gintis. Oxford: Oxford University Press.
Cameron, Lisa and Manisha Shah. 2014. Can Mistargeting Destroy Social Capital and Stimulate Crime? Evidence from a Cash Transfer Program in Indonesia. Economic Development and Cultural Change, Vol. 62, No. 2 (January), pp. 381-415
Candelo, Natalia and Sandra Polania-Reyes (2008) Pasos metodológicos de un diseño experimental para medir capital social y acción colectiva en seis ciudades latinoamericanas (2008) [Methodological steps for an experimental design to measure social capital and collective action in six Latinamerican cities] Working paper CEDE (17) Universidad de los Andes.
C.M. Capra, K. Lanier, S. Meer. 2008. Attitudinal and behavioral measures of trust: a new comparison. Working paper Emory University. Retrieved at: https://www.gate.cnrs.fr/IMG/pdf/Capra.pdf
Cardenas, Juan Camilo. 2003. "Real Wealth and Experimental Cooperation." Journal of Development Economics, 70, pp. 263-89.
Cardenas, Juan Camilo, Alberto Chong, Hugo Ñopo, Andrew W. Horowitz, and Daniel Lederman. "To What Extent Do Latin Americans Trust, Reciprocate, and Cooperate? Evidence from Experiments in Six Latin American Countries [with Comments]." Economia (2009): (2) 45-94.
Cárdenas, Juan Camilo, Alberto Chong, and Hugo Ñopo. "Stated social behavior and revealed actions: Evidence from six Latin American countries." Journal of Development Economics 104 (2013): 16-33.
Cárdenas, Juan Camilo and Christian Jaramillo. 2007. "Cooperation in Large Networks: An Experimental," In Documentos CEDE, Universidad de los Andes-Cede. Bogota.
Carpenter, Jeffrey. 2002. "Measuring Social Capital: Adding Field Experimental Methods to the Analytical Toolbox," In Social Capital, Economic Development and the Environment, ed. S. Ramaswamy, J. Isham and T. Kelly. Edward Elgar.

Carpenter, Jeffrey; Amrita Daniere and Lois Takahashi. 2004. "Cooperation, Trust and Social Capital in Southeast Asian Urban Slums." Journal of Economic Behavior and Organization, 55(4), pp. 533-52.

Casey, Katherine, Rachel Glennerster, and Edward Miguel. 2012 Reshaping Institutions: Evidence on Aid Impacts Using a Preanalysis Plan. Quarterly Journal of Economics. 127 (4): 1755-1812 first published online September 7, 2012 doi:10.1093/qje/qje027

Chattopadhyay, Raghabendra, and Esther Duflo. "Women as policy makers: Evidence from a randomized policy experiment in India." *Econometrica* 72.5 (2004): 1409-1443.

Christoforou, A., & Davis, J. B. (Eds.). (2014). Social Capital and Economics: Social Values, Power, and Social Identity. Routledge.

Coleman, J. (1988). Social capital in the creation of human capital. American Journal of Sociology, 94, S95–S120.

Coleman, James Samuel. 1990. Foundations of Social Theory. Cambridge, Massachusetts and London, England: Harvard University Press.

Cohen E, Franco R, eds. 2006. Transferencias con Corresponsabilidad. Una Mirada Latinoamericana. México: Fac. Ciencias Soc. (FLACSO). 434 pp.

Croson, Rachel T. A. 1996. "Contributions to Public Goods: Altruism or Reciprocity?," In. The Wharton School, University of Pennsylvania.

De La O., Ana 2013 "Do conditional cash transfers affect electoral behavior? Evidence from a randomized experiment in Mexico." American Journal of Political Science 57, no. 1: 1-14.

DNP. 2010. The journey: ten years of Familias en Acción [El camino recorrido: Diez Años de Familias en Acción]. Bogota, Colombia; July; Acción Social Departamento Nacional de Planeación.

ECLAC United Nations Economic Commission for Latin America and the Caribbean. *Aprender de la experiencia: el capital social en la superación de la pobreza*. Edited by Irma Arriagada. Vol. 86. United Nations Publications, 2005.

Fearon, J., Humphreys, M. & Weinstein, J. M. 2009, "Can Development Aid Contribute to Social Cohesion after Civil War? Evidence from a Field Experiment in Post-Conflict Liberia", American Economic Review, Vol. 99, No. 2

Fiszbein, Ariel; Norbert. R. Schady; Francisco H.G. Ferreira; Margaret Grosh; Nial Kelleher; Pedro Olinto and Emmanuel Skoufias. 2009. Conditional Cash Transfers: Reducing Present and Future Poverty. World Bank Publications.

Gächter, Simon, Herrmann, Benedikt, Thoni, Christian, 2004. "Trust, voluntary cooperation, and socio-economic background: survey and experimental evidence," Journal of Economic Behavior & Organization, Elsevier, vol. 55(4), pages 505-531, December.

Gilligan, Michael J., Benjamin J. Pasquale, and Cyrus Samii. 2014 "Civil War and Social Cohesion: Lab-in-the-Field Evidence from Nepal." *American Journal of Political Science*. 58, Issue 3, p.604–619.

Grossman S. J. and O. Hart (1980). “Takeover Bids, The Free-Rider Problem, and the Theory of the Corporation.” *The Bell Journal of Economics*, Vol. 11, No. 1 (Spring, 1980), pp. 42-64.

Jakiela, Pamela, Edward Miguel, and Vera L. te Velde. "You’ve earned it: estimating the impact of human capital on social preferences." Experimental Economics (2014): 1-23.

Jones, Benjamin F., and Benjamin A. Olken. "Do leaders matter? National leadership and growth since World War II." *The Quarterly Journal of Economics* (2005): 835-864.

Karlan, Dean. 2005. "Using Experimental Economics to Measure Social Capital and Predict Financial Decisions." American Economic Review, 95(5), pp. 1688-99.

Labonne, Julien, The Local Electoral Impacts of Conditional Cash Transfers: Evidence from a Field Experiment," Journal of Development Economics, September 2013, 104, 73-88.

Latham, Gary P., and Lise M. Saari. "Importance of supportive relationships in goal setting." Journal of Applied Psychology 64, no. 2 (1979): 151.
Latorre López, María Carolina. 2004. "Sobre la relación positiva entre el capital social y la violencia urbana: un análisis teórico y empírico," [About the positive relationship between social capital and urban violence: a theoretical and empirical analysis]. In Documentos CEDE no. 36. Bogota DC.
Ledyard, J. O. 1995. "Public Goods: A Survey of Experimental Research," In The Handbook of Experimental Economics, ed. A. E. Roth and J. Kagel, 111-94. Princeton, NJ: Princeton University Press.
MacAuslan, Ian, and Nils Riemenschneider. 2011 "Richer but resented: What do cash transfers do to social relations?" IDS Bulletin 42, no. 6: 60-66.
http://onlinelibrary.wiley.com/doi/10.1111/j.1759-5436.2011.00274.x/pdf
Macours, Karen and Vakis, Renos. 2014 "Changing Households' Investment Behaviour through Social Interactions with Local Leaders: Evidence from a Randomised Transfer Program". The Economic Journal, 124: 607–633. doi: 10.1111/ecoj.12145
Manacorda, Marco, Edward Miguel, and Andrea Vigorito. 2011. "Government Transfers and Political Support." American Economic Journal: Applied Economics, 3(3): 1-28.
Mansuri, Ghazala, and Vijayendra Rao. Localizing development: does participation work?. World Bank Publications, 2012.
MESEP, 2012. Mission for Splicing Series Employment, Poverty and Inequality (MESEP). Monetary poverty in Colombia: New methodology and figures 2002-2010 [Misión para el Empalme de las Series de Empleo, Pobreza y Desigualdad (mesep). Pobreza monetaria en Colombia: Nueva metodología y cifras 2002-2010] Report. DNP, DANE. Bogota, Colombia.
Nguyen, Tu Chi, and Matthias Rieger. "Community-driven development and social capital: Evidence from Morocco." (2014). EUI Working Paper MWP 2014/02
Nupia, Oskar. "Anti-poverty programs and presidential election outcomes: Familias en acción in colombia." Universidad de los Andes (2012). Revised version of Documento CEDE 14 (2011).
Ñopo, Hugo, Juan Camilo Cárdenas, Sandra Polania-Reyes, Natalia Candelo, Sebastian Calónico, Alberto Chong. 2008. To What Extent Do Latin Americans Trust and Cooperate? Field Experiments on Social Exclusion in Six Latin American Countries In: IADB IPES 2008 Report: Outsiders? The Changing Patterns of Exclusion in Latin America and the Caribbean (David Rockefeller/Inter-American Development Bank). Chapter 7.
Olson (1965). *The logic of collective action*. Cambridge: Harvard University Press
Ostrom, E. (1998). "A behavioral approach to the rational choice theory of collective action." *The American Political Science Review*, 92 (1): 1-22.
Pellerano, L. 2008, 'Assessing the impact of conditional cash transfer programs on social participation: evidence from Colombia', Institute for Fiscal Studies, Manuscript.
Pellerano, Luca, Marta Moratti, Maja Jakobsen, Matěj Bajgar and Valentina Barca, 2014. Child Grants Program Impact Evaluation. Follow-up Report. April.
www.fao.org/fileadmin/user_upload/p2p/Documents/draft_CGP_Follow_Up_v11_out.pdf
Polania-Reyes, Sandra. 2005. "Capital Social E Ingreso De Los Hogares Del Sector Urbano En Colombia." [Social Capital and Household Income in Urban Colombia]. Revista Desarrollo y Sociedad - Universidad de los Andes, 50(2).
Pop-Eleches, Cristian, and Grigore Pop-Eleches. "Targeted Government Spending and Political Preferences." Quarterly Journal of Political Science 7, no. 3 (2012): 285-320.
Putnam, Robert D. 1993. Making Democracy Work: Civic Traditions in Modern Italy. Princeton, N.J.: Princeton University Press.

Putnam, Robert D. "Bowling alone: America's declining social capital." *Journal of Democracy* 6 (1995): 68.
Samuelson, Paul A. (1954). "The Pure Theory of Public Expenditure." *Review of Economics and Statistics* 36 (4): 387–389.
Tajfel, Henri. "Social psychology of intergroup relations." Annual review of psychology 33, no. 1 (1982): 1-39.
Uphoff, N. (1999), "Understanding social capital: learning from the analysis and experience of participation", Social capital. A multifaceted perspective, P. Dasgupta y I. Serageldin (eds.), Washington, D.C., Banco Mundial.
Warren, Mark R. "Community Building and Political Power A Community Organizing Approach to Democratic Renewal." American Behavioral Scientist42, no. 1 (1998): 78-92.
Wong, Susan. "What have been the impacts of World Bank Community-Driven Development Programs? CDD impact evaluation review and operational and research implications." World Bank, Washington, DC (2012).
Zárate, Román A., Emily Conover, Adriana Camacho, and Javier E. Báez. "Conditional Cash Transfers, Political Participation, and Voting Behavior." 2013. http://cie.itam.mx/englishversion/faculty/fea_elect_Adriana%20Camacho.pdf
Zucco, Cesar, 2011. Cash-Transfers and Voting Behavior: An assessment of the political impacts of the Bolsa Familia Progam, mimeo, Princeton University.

Appendix A. Experimental Design and Procedure 2007 and 2008

i. About Cartagena, Pozón, Ciénaga and Familias en Acción (FA)

Cartagena is the fifth largest city in Colombia, with 993 thousand inhabitants in 2008. It is the third poorest city in the country, with 40.2% poor and 6.9% in extreme poverty in 2008 (MESEP, 2012). *Ciénaga* and *Pozón* belong to the poorest locality (i.e. the lowest level of income, the lowest education coverage, the highest infant mortality rates and the worst living conditions in Cartagena (CCV (2011)). In 2009, Pozón is recognized as the densest neighborhood with an area of 273 Ha and 45 thousand inhabitants while Ciénaga has 463 Ha and 102 thousand inhabitants (see Figure A1)[70]. In 2006, Pozón and Ciénaga are considered by the local authorities as very similar, with a percentage of households with lower income (56%) and the same average time in school (6 years)[71].

The pilot in Pozón was implemented in January 2005 with 5 thousand Sisben 1 households. A new enrolment wave took place in March 2006 for 2.5 thousand displaced households (i.e. households that were forced to leave their home because of the civil conflict). After that, displaced households have been allowed to enroll in the program at any time. In late August 2007, 35.5 thousand households were enrolled in Cartagena, including households from Pozón. In our data set 58 of 404 individuals who attended the follow-up (14.4% of Pozón sample) were enrolled in the program in 2007. The following enrollment waves after 2007 (for non-displaced households) took place in 2009 (32.000 households), in 2012 (22,000 households) and 2013 (7,000 households). Regarding the payment procedure of FA, the first payment in Pozón was in March 2005 followed by a bi-monthly payment. In 2007, the first payment in Ciénaga was in October 2007.

[70] Source: CCV. URL: http://www.cartagenacomovamos.org/ucg.swf
[71] Calculations made with data provided by the Mayor office in Cartagena in the study by Pérez et al (2007).

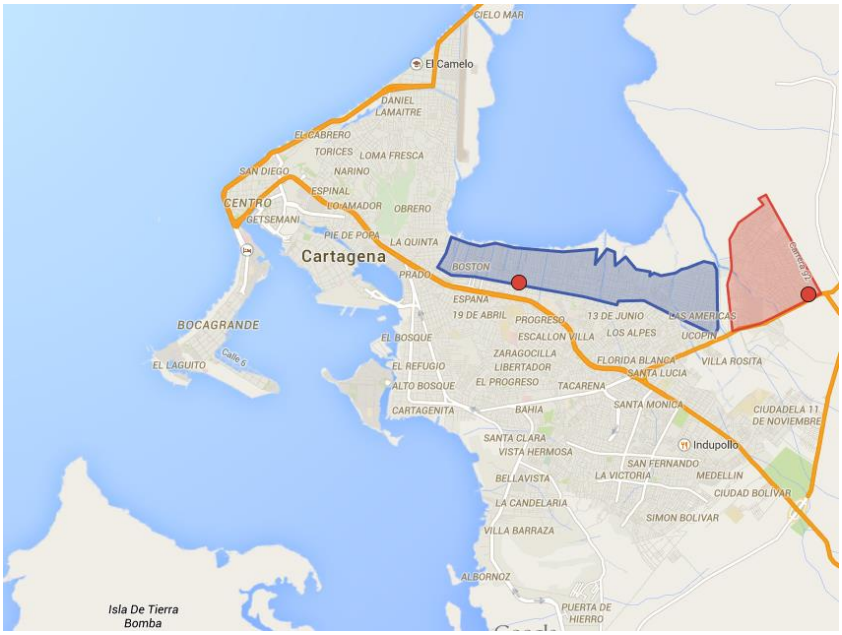


Figure A1. Cartagena, Pozón and Ciénaga
Source: http://midas.cartagena.gov.co/ and Map data © 2014 Google. Red dots are where the sessions were held. The red area is Pozón and the blue one Ciénaga.

At the EC, ML and beneficiaries follow up on their current health and education status. They discuss any aspect related to the community, and by doing so they reach a common ground to make decisions and take actions aimed at improving their life conditions. It is also a space for enjoyment among peers. From January 2005, ECs were held quarterly. However, a ML was allowed to organize EC with her beneficiaries whenever she considered. The number of ECs was determined by how proactivity the ML was. There were differences in the EC in Pozón between the period 2005 to 2007 and from 2007 to 2008. As the ML were trained, they felt empowered within their community, displacing other community leaders.

In addition to the EC, the beneficiaries take part in the general assembly. The general assembly is a public meeting where beneficiaries discuss and decide about problems affecting beneficiaries in Cartagena. The ML are elected in the general assembly. There are four annual assemblies, taking place on a date set by the local office. Although the national office does not make attendance to the assemblies a mandatory requirement, from 2005 to 2010 the local office made it so.

The percentage of the neighborhood population receiving the program was 79% in Pozón in 2006 and 22.4% in all of Cartagena in 2008. This CCT is targeted to women, like every other CCT in Latin America (i.e. Mexico, Nicaragua, Argentina, Chile, Paraguay, and Brazil. See Molyneux, 2011).

ii. Recruitment and sampling

In 2007, we had planned to run 15 sessions with 375 participants in each neighborhood. The attendance rate was 89.07% and 91.2% in Ciénaga and Pozón respectively, given that one session had to be cancelled in each neighborhood and in some sessions there were less than 25 attendees.

In 2008, we sent 500 invitations to individuals who had not previously participated into the experiment and that were beneficiaries of *Familas en Acción*. We assumed a response rate of 70% and expected to run 14 sessions with 350 attendees. The actual attendance rates for the new participants was 74.2% and 80.8% in Ciénaga and Pozón respectively. This means we conducted additional sessions with 75 additional new participants.

In addition to the new participants we also invited individuals who had participated in 2007. We planned to conduce 13 sessions (rather than the original 14). However, the attrition rate was 10.77% and 27.08% in Ciénaga and Pozón respectively, caused mostly by the fact that we had only three working days in a specific week and the leaders in charge of the recruitment explained most of the participants in 2007 who could not attend in 2008 were working or had other commitments.

iii. Experimental Design: the Voluntary Contribution Mechanism[72]

Each player receives an endowment of one token to be invested either in a private or a group account. The decision is made privately and simultaneously. The earnings are calculated in the following way: if the player chooses to invest in the private account, the token is converted into $5 and will be given entirely to her. In addition each player receives, regardless of how she has invested her own her token, $0.40 for each token invested in the group account by any other member in the group of 25 players. Therefore, her total earnings at the end of this round are ($5) + ($0.40 x Sum of Tokens invested by the group). If the player chooses to invest her token in the group account, she will receive 0.4 for each token invested in the group account by her and in the rest of the group. In this case her total earnings at the end of the round will be ($0) + ($0.40 x Sum of Tokens invested by the group). Each player makes her private decision by selecting a card which says if she is going to invest her money in the group account or to keep it for herself (i.e. private account). The experimenter then collects the "decisions cards," totals them up, multiplies by $0.40 the amount and credits the relevant amounts to each player. The relevant amounts, however, are only revealed and paid at the end of the session and after a second round of the same game.

In the first round, each player has to decide where to invest her token. The second round is a repetition of the first, except that the players are allowed to discuss for ten minutes before making simultaneously their private, anonymous decision. Communication is completely unstructured and during the discussion, the players can talk about whatever they want but they cannot leave the room. No one, except the experimenter, knows the other players' contributions in the first round.

The players do not even know their own payout in the first round when they discuss with the other players or when they play the second round. At the end of the session, however, the totals but not the individual strategies are announced. The players are paid in cash privately at the end of the session after taking a snack. The payoff resulting from both rounds are paid together with a show-up fee of $2 to cover the transportation costs of each participant.

This experiment has been designed to set up a situation for 25 people where there is a social dilemma, illustrated in Figure A2. This figure shows the two possible individual outcomes

[72] The experimental design of the VCM described here was developed by Juan Camilo Cárdenas, Maria Claudia Lopez, Natalia Candelo and Sandra Polanía-Reyes.

(playing public or private) as a function of the number of individuals that contribute to the group account. Clearly, the dominant strategy for the j-1 player is to choose not to invest the token in the group account because each token contributed yields only $0.40 to its contributor, no matter what the others do. Therefore, each player would want to "free ride" on the others the benefits of the group account and the Nash equilibrium is that everyone invests in the private account (point A). In this case, the group earnings would be $25 (players) x $5 (private account return) = $125. The social optimum where the group would be best off is that everyone contributes (point B) and each token contributed yields $10 (25 (players)*0.4 (group account return) to the others at no cost to them. In this case, the group earnings would be 25 (players) x $10 (individual benefits of the group account) = $250. Notice that it is necessary that at least 12 players contribute to the public account for a contributor to this account to earn the same as in the Nash equilibrium, where everybody invests her token to the private account. With smaller groups, critical mass is more difficult to achieve.

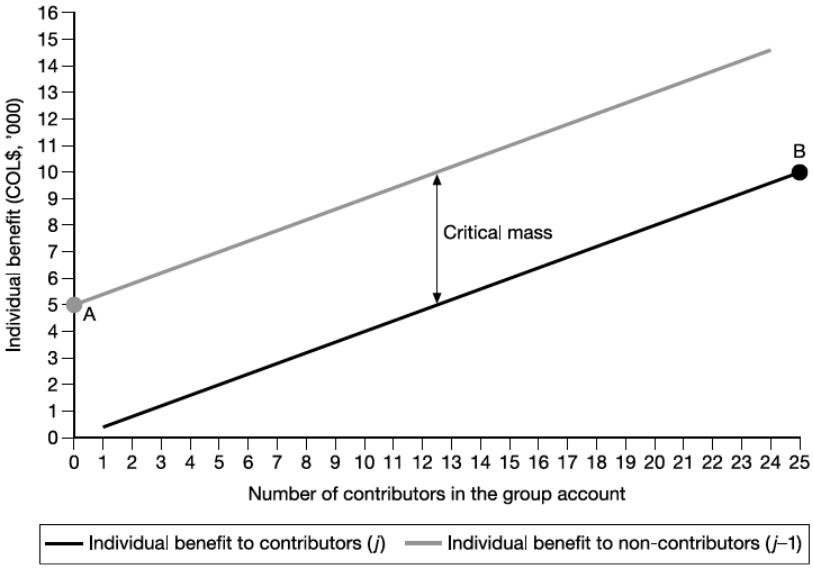


Figure A2. The basic structure of the social dilemma in the VCM
Source: Attanasio et al. (2009)

In 2008, in addition to (and after) the public goods game we describe in what follows, we also played a 'minimum effort coordination game', which could measure 'social capital' as the ability to coordinate on the best Nash equilibrium. As this game was only played in the follow up when the program was operating in both neighborhoods, we cannot use it to evaluate the impact of FA on social capital. We therefore do not discuss its details in this article.

iv. Post experimental data

In the survey we collect at the end of the session, we ask questions that have been suggested as possible measures of social capital such as membership in organizations and trust. This survey yields some insights into the individual characteristics that could affect the decision of contribution (i.e. volunteerism, participation in community activities, leadership and attendance to their meetings) and gives also information to examine the external validity of our results.

The survey is completed by a module that aims at filling in a 'networks map' for each session, asking the subjects questions about their relationship to other participants. In

particular, in the networks questionnaire, we asked to each participant about her relationship with every other player in the group ("how do you consider person Y?") in three categories: relatives, friends and acquaintances. The network information allows us to measure the degree of connectivity among the players in each session[73]. We also ask whether the player considers any other player as a leader in the group (only one person could be named). This variable provides us with information about the presence of leaders and is useful to determine their potential role in the game outcome.

The network structure determines the degree of cohesion in a group and the ability to overcome the costs of collective action and obtain benefits related with cost of transaction, information and risk management. In the VCM experiment, the structure affects the player's willingness to communicate with others and the ability to do so in an effective and organized way to obtain the maximum social outcome. In addition, the structure of the social network can determine individual perception of group's characteristics and network's quality (Cárdenas and Jaramillo, 2007); if there is no social network (i.e. all participants are strangers) the cost of effective communication could be higher.

v. Outcome of the VCM

Table A1. Willingness to cooperate in 2007 and 2008 in the first round

	2007				2008			
	All	**Pozón**	**Ciénaga**	**Difference**	**All**	**Pozón**	**Ciénaga**	**Difference**
Median percentage of contributors	3.6	7.0	0.0	27.3	9.8	0	19.0	-19.0*
	(3.6)	(7)	(0)	(0.65)	(5.5)	(0)	(10.3)	(10.1)
Maximum percentage of contributors	85.2	100	70.1	29.9**	89.5	85.5	93.2	-7.7
	(6.9)	(0.0)	(12.9)	(12.7)	(5.8)	(9.8)	(6.8)	(11.7)
Observations	28	14	14	-	53	26	27	-

Standard errors that are clustered at the session level in parenthesis. The standard errors for the median and maximum statistics are calculated at session level. * Significant at 10%;** significant at 5%; *** significant at 1%.

vi. Comparing different measures

Given the three sets of measures we have described, an interesting question is the extent to which they co-vary. We conclude this section with such analysis, which we conduct at the individual level and at session level. We explore the association between individual cooperative behavior in the first round of the VCM game and each of our possible measures of social capital at the session level. For any given social capital measure, Table A2 shows the coefficients obtained in a number of probit regressions that relate the probability that an individual contributes to the public good in the first round to each of the measures of social capital considered in Table 4 and Table 5 at the individual level. The purpose of these tables is simply to illustrate the association between the behavior in the VCM game we observe and other (self-reported) forms of behavior that have been used in the literature as indicators for pro-social behavior.

73 We are aware that asking the subjects about their acquaintance's trustworthiness would have given us a more accurate measure of social capital. There were two reasons we got only a measure for connectivity. First, we replicated the same experimental and survey design which was conducted in the FA's first and second follow-up evaluations in rural areas. Second, we were constrained by time logistics. However, we consider collecting a trustworthiness measure in future studies.

In Table A2 we report three coefficients for each variable: that estimated using baseline data, the one estimated with follow up data (using only 'new' players) and the one obtained pooling all observations. For the baseline, the coefficients on all measures of social capital have the expected sign and most of them are significantly different from zero. For instance, a participant is less likely to contribute to the public good in the game if she does not participate in community decisions or meetings and if she perceives a selfish attitude in the community. Individuals who consider no one can be trusted are more likely to not to cooperate. However, in the follow up, no individual measure is significantly associated with our measure of pro-social behavior. It seems that the relationship between our measure of social capital and several of the variables traditionally used is much harder to capture in the follow up. As with the follow up, when we pool the data, we are unable to detect any significant association between our variable and the traditional measure of social capital.

Table A2. Social capital measures at individual level and individual cooperation

Dependent variable: 1 if player contributed to the group account in the first round					
Probit marginal effects			**2007**	**2008**	**Pooled**
Civic Participation	Membership in at least one organization		0.030	0.008	0.007
			(0.032)	(0.036)	(0.025)
	Participation in neighborhood decisions		0.082***	-0.014	0.020
			(0.031)	(0.039)	(0.025)
	Participation in the neighborhood meetings		0.056**	-0.034	-0.005
			(0.029)	(0.033)	(0.024)
Voting behavior	Vote Local Elections (2003 and 2007)		0.026	-0.002	0.032
			(0.033)	(0.044)	(0.028)
	Vote Presidential Elections (2006)		0.077***	-0.012	0.049*
			(0.026)	(0.052)	(0.029)
Trust and cooperation perception	Trust	Most people	0.032	-0.048	-0.010
			(0.061)	(0.090)	(0.055)
		Few people	0.069***	0.016	0.042*
			(0.024)	(0.035)	(0.021)
		None	-0.085***	-0.004	-0.043*
			(0.028)	(0.037)	(0.024)
	Perception within the community	Cooperation	0.045	-0.021	0.014
			(0.033)	(0.031)	(0.023)
		Reciprocity	0.082**	-0.004	0.030
			(0.037)	(0.047)	(0.031)
		Selfishness	-0.095***	0.021	-0.030
			(0.035)	(0.026)	(0.023)
Individual Social Integration	Number of player's relatives		-0.162**	0.017	-0.051
			(0.067)	(0.046)	(0.035)
	Number of player's friends		-0.000	0.008	0.006
			(0.009)	(0.011)	(0.007)
	Number of player's acquaintances		0.004**	-0.013	-0.001
			(0.002)	(0.016)	(0.003)
	Number of player's connections		0.003*	0.003	-0.001
			(0.002)	(0.010)	(0.003)
Player is identified as leader in the session			-0.014	-0.026	-0.037
			(0.029)	(0.045)	(0.026)
Observations			676	714	1390

* Significant at 10%; ** significant at 5%; *** significant at 1%. Marginal effects reported. Robust standard errors clustered by session in parenthesis after probit estimation.

In addition to looking at individual behavior, it might be useful to complement the evidence in Table A2 by looking at different measures of social capital in *groups.* Our context offers a natural definition of a group, as the game session. While a session is an artificial construct

generated by the administration of the experiment, considering session means can be informative about the performance of different measures to the extent to which individual behaviors and idiosyncrasies averaged out.

In Table A3, we report, for both years and for the pooled data, the Pearson correlation coefficients across sessions, between the percentage of individuals contributing to the public good in the first round, and the session averages of the various measures we have considered in Tables 4 and 5.

Table A3. Spearman Correlation coefficients at the session level

Percentage			2007	2008	Pooled
Civic Participation	Membership in at least one organization		0.57***	0.33*	0.29**
	Participation in neighborhood decisions		0.7***	0.10	0.3**
	Participation in the neighborhoods meetings		0.62***	0.08	0.17
Voting behavior	Vote Local Elections (2003 and 2007)		0.39**	0.24	0.37***
	Vote Presidential Elections (2006)		0.64***	-0.15	0.36***
Trust and cooperation perception	Trust	Most people	0.21	0.11	0.14
		A few people	0.52***	-0.24	0.13
		Nobody	-0.72***	0.16	-0.25*
	Perception within the community	Cooperation	0.48***	-0.07	0.23*
		Reciprocity	0.34*	0.30	0.13
		Selfishness	-0.6***	-0.19	-0.36***
Individual Social Integration	Average number of relatives		-0.64***	0.33*	-0.18
	Average number of friends		0.14	0.25	0.24*
	Average number of acquaintances		0.34*	0.29	-0.04
	Average number of connections		0.37*	0.29	-0.04
Network features	Relatives density [a]		-0.66***	0.23	-0.13
	Friendship density		0.16	-0.03	0.16
	Acquaintances density		0.39**	0.16	-0.07
	Connections density		0.41**	0.1	-0.09
Percentage of players identified as leader in the session			0.62***	0.24	0.22*
Observations			28	29	57

[a] Network density is calculated as the number of connections /(number of connected people)*(number connected people -1). For example, Friends density is calculated as the number of friends/(number of people with friends)*(number of people with friends -1). Robust standard errors clustered by session. * Significant at 10%; ** significant at 5%; *** significant at 1%.

Interestingly, the pattern of correlations of contributions rates in the VCM game and the prevalence of friendships in the sessions is not only positive and significant but does not change much over time. The measure based on connections, however, changes sign in 2008. When we pool the data, in the last column, we find a pattern which is very similar to the one observed in the baseline.

The evidence is mixed. In 2007, all the 'traditional' social capital measures are significantly associated with cooperation in the game and with the expected sign. For instance, in sessions in which many participants perceive their community as populated by selfish individuals and report that nobody can be trusted, contributions rates to the public good are low, which explains the first two negative signs in the first column of the Table A3. However, this evidence does not seem completely robust. In 2008, while the relationship between measures based on civic participation is still positive and significant, for the first two variables considered, the correlation coefficient is much lower. As for the measures based on voting behavior, the one based on local elections is still positive, significantly different from zero and roughly of the same order of magnitude. The measure based on voting in the presidential elections, however, is now negatively related to contribution rates to the public good. It is similar for the measures

based on answers to the 'trust' questions. The correlation with the first and third indicator (trusting most people and trusting nobody) is now not significantly different from zero and the one with the middle one (trusting a few people in the community) changes sign. As for the measure based on perceptions on cooperation, the correlation coefficient changes sign, while the correlation coefficients with the measures based on the perception of reciprocity and selfishness stay the same.

Overall, the results presented in Table 7 are consistent with the evidence presented in Table A2 and Table A3. While the traditional measures deliver similar results in terms of impact, the evidence is much less precise and, often, the results are not statistically different from zero. The use of our social capital behavioral measure instead affords us much more precise estimates.

Appendix B. Time Trends in Structural Characteristics across groups

Table B1. Time trends in structural characteristics across neighborhoods – independent cross section

		Pozón		Ciénaga	
		Difference	**Std. errors**	**Difference**	**Std. errors**
General characteristics	Percentage of female participants	-1.2*	0.6	0.2	1.2
	Average age (years)	-1.8**	0.8	2.7***	0.8
	Number of years living in the neighborhood	0.1	0.6	-0.1	0.8
	Percentage displaced	-2.5	3.2	1.9	1.9
	Percentage household head	2.9	2.8	10.8**	4.5
	Percentage Single	5.1**	2.1	0.3	3.1
	Percentage married of civil partnership	0.3	3.0	4.5	4.9
Educational level (percentage)	None (level 0)	0.3	1.2	-1.5	1.6
	Primary incomplete (level 1)	3.0	3.9	2.4	3.3
	Primary complete (level 2)	1.6	2.8	-0.2	2.6
	Secondary incomplete (level 3)	-2.2	3.6	0.5	3.9
	Secondary complete (level 4)	-5.8*	3.5	-3.7	3.5
	More than secondary complete (level 5)	3.1	2.8	2.5	1.9
Dwelling characteristics	Household size	-0.0	0.1	-0.1	0.2
	Number of people per room	0.4**	0.1	0.5***	0.1
	Percentage with dirt floor	6.4**	3.0	-15.2***	4.2
	Percentage owning own house	-16.3***	2.9	-6.8*	4.0
Public services (percentage)	Water by pipe	-2.8	1.8	6.5	4.7
	Sewer system	3.0	4.2	13.0***	4.6
	No electricity	3.2**	1.6	-3.5**	1.6
	Land phone	-1.6	2.3	-0.3	3.2
Observations		688		702	

Robust standard errors, clustered by session. *, **, *** Significant at 10%, 5% and 1% resp. Observations in follow-up are independent sessions. In Pozón the number of observations is 342 (resp. 334 Ciénaga) in the baseline and 346 (resp. 368 Ciénaga) in the follow-up.

Table B2. Time trends in structural characteristics across neighborhoods 2007 and 2008 – panel data

		Pozón		Ciénaga	
		Difference	**Std. errors**	**Difference**	**Std. errors**
General characteristics	Percentage of female participants	-0.4	0.4	0.0	1.1
	Average age (years)	0.4	0.8	0.9	0.8
	Number of years living in the neighborhood	0.8	0.8	0.0	0.8
	Percentage displaced	0.0	3.4	0.0	2.1
	Percentage household head	11.4***	3.8	2.8	4.4
	Percentage single	0.0	2.2	3.4	2.5
	Percentage married of civil partnership	-5.1	3.2	4.1	4.8
Educational level (percentage)	None (level 0)	0.0	1.2	-1.7	1.6
	Primary incomplete (level 1)	0.4	4.1	-0.3	3.5
	Primary complete (level 2)	-2.5	3.0	-2.1	2.5
	Secondary incomplete (level 3)	3.4	4.0	3.4	4.3
	Secondary complete (level 4)	-1.7	4.0	1.4	3.9
	More than secondary complete (level 5)	0.4	1.4	-0.7	2.4
Dwelling characteristics	Household size	-0.1	0.2	-0.6***	0.2
	Number of people per room	0.1	0.1	0.2	0.1
	Percentage with dirt floor	-3.0	4.3	-2.8	3.9
	Percentage owning own house	-6.8*	3.9	-5.2	4.0
Public services (percentage)	Water by pipe	-0.4	1.6	8.6*	4.9
	Sewer system	13.9***	4.3	13.1***	4.4
	No electricity	-1.3	1.0	-3.4*	2.0
	Land phone	-3.8*	0.0	-5.2*	3.1
Observations		474		580	

Robust standard errors, clustered by session. *, **, *** Significant at 10%, 5% and 1% resp. 237 individuals in Pozón (resp. 290 Ciénaga).

Appendix C. Pozón and Ciénaga 2007

i. First approach to identify the effects of FA on willingness to cooperate to a public good

The comparison between *Pozón* and *Ciénaga* can be affected, however by differences between the two neighborhoods other than the operation of the program. Table B1 shows that these differences are present and important. If 'social capital' is a 'luxury', the fact that participants from *Pozón* are marginally less poor could explain why they show a larger propensity to contribute to the public good.

One first possibility to address this issue is to control for observable variables that capture differences between the two neighborhoods (but are not affected by the program). The reduced form equation to be estimated is the following:

(1) $$Y_{is} = \alpha + \beta X_{is} + \gamma D_i + \theta S_s + v_s + \varepsilon_{is}$$

Where Y_{is} is the individual decision to contribute to the group account in session s in 2007; D_i is a dummy for treatment; X_i are individual observable characteristics, S_s are session level characteristics and ε_{is} are i.i.d. Gaussian distributed error terms with mean zero and variance $\sigma_\varepsilon^2 = 1$, independently of v_s, which are iid, $N(0, \sigma_v^2)$.

In Table C1 we show the impact of being a player from *Pozón* on the probability to contribute to the group account in 2007, after controlling for a variety of observable variables. All four columns show results obtained from a probit regression. In Column (i), we

do not add any controls. In column (ii) we control only for individual characteristics, such as the age and the education level of the participant (the complete set of results is shown in Table C2). The introduction of these variables does not affect much the size of the measured impact, which remains large and significantly different from zero.

Table C1. Program's Impact on the Cooperative Decision - First Round in 2007

Dependent variable: 1 whether player contributed to the group account in the first round

Independent Variables	(i)	(ii)	(iii)	(iv)
Dummy *El Pozón*	0.265***	0.277***	0.277***	0.223***
	(0.044)	(0.071)	(0.051)	(0.048)
Basic controls - participant characteristics	No	Yes	Yes	Yes
Session characteristics	No	No	Yes	Yes
Session composition (info within session)	No	No	No	Yes
Observations	676	670	670	670

*Significant at 10%; **significant at 5%; *** significant at 1%. Marginal effects reported. Robust standard errors, clustered by session, in parenthesis after probit estimation.

While in specification (ii) we control only for individual and household characteristics (such as age, education and so on), in Column (iii), we control for session characteristics. An issue that worried us considerably on the field was the possibility that individuals who played in early sessions would 'contaminate' other individuals that were about to enter subsequent session by talking to them and commenting on the game. While we tried to avoid these contacts as much as possible and instructed the subjects not to talk to subsequent players, some contacts were unavoidable. For this reason, we control for the sequence order in which a particular session is played in a day and also for results in previous sessions[74]. We do find these session effects to be significant. However, their introduction does not change the size of the marginal impact, partly because these effects were present in both treatment and control sessions.

In Column (iv) we estimate equation (1) and control for session composition effects. In particular, we control for session heterogeneity with variables such as the deviation of standard deviation of level of education in the session with respect to the mean of the neighborhood and percentage of participants in the session with less than secondary incomplete. Controlling for these variables lowers the point estimates of the treatment effect substantially. However, the coefficient is still significantly different from zero and of a sizeable magnitude.

The results in Table C1 can still be misleading. In the same way in which the two neighborhoods can be systematically different in terms of observable variables, there might be differences in terms of unobservable variables. Such differences would bias the estimates of the impact of the program derived from comparisons that control only for observable variables. This is why the main specification we consider is a difference-in-differences regression.

[74] Particularly, in order to control for contiguous sessions correlation, we include in specification (iv): a) a dummy for the first session each day; b) a variable capturing the deviation from the neighborhood mean of the average contribution to the public account in the previous 2 sessions, and c) the order in which the session came about in the full sequence of sessions within the neighborhood.

Table C2. Controls for the decision to contribute in the group account in the first round 2007-Table C1

Dependent variable: 1 if player contributed to the group account in the first round

	Independent Variables	(ii)	(iii)	(iv)
Basic Controls - Participant's characteristics	If the player has a partner	0.065*	0.057	0.057*
		(0.034)	(0.035)	(0.034)
	Level of education (0 to 5)	0.020*	0.019*	0.019*
		(0.011)	(0.010)	(0.010)
	Household size	0.015**	0.012	0.014*
		(0.007)	(0.008)	(0.008)
	Ground floor (house)	-0.041	-0.033	-0.034
		(0.030)	(0.028)	(0.029)
	Number of years living in the neighborhood	0.001	0.000	0.000
		(0.002)	(0.002)	(0.002)
	1 if the player's home has water pipe access	0.125***	0.113***	0.113***
		(0.030)	(0.025)	(0.024)
	1 if the player's home has Sewage	-0.040	-0.035	-0.031
		(0.035)	(0.031)	(0.031)
	1 if the player's home has no electricity	0.127	0.094	0.123
		(0.118)	(0.117)	(0.129)
	1 if the player is the head of household	0.043	0.039	0.050
		(0.052)	(0.051)	(0.053)
	Age	0.002	0.002	0.002
		(0.002)	(0.002)	(0.002)
	1 if the player is unemployed	0.046	0.043	0.041
		(0.073)	(0.069)	(0.070)
	1if the player has her own housing	-0.048	-0.057	-0.057
		(0.043)	(0.043)	(0.043)
	1 if player belongs to the fifth quintile per capita income	-0.011	-0.004	-0.005
		(0.033)	(0.031)	(0.031)
	1 if the player has a phone	0.071	0.071	0.071
		(0.057)	(0.054)	(0.054)
	Number of rooms in the house	-0.018	-0.012	-0.012
		(0.018)	(0.018)	(0.018)
	1 if the player is displaced (self-declared)	0.037	0.017	0.011
		(0.037)	(0.032)	(0.034)
	1 if the player declares she understood everything	-0.010	-0.021	-0.017
		(0.046)	(0.044)	(0.041)
	1 if the household receives other support from another institution	-0.056	-0.036	-0.031
		(0.036)	(0.032)	(0.033)
Session Characteristics	Number of players in session		-0.012	-0.023***
			(0.008)	(0.006)
	1 if there is at least one man in the session		0.272***	0.245***
			(0.056)	(0.044)
	Experimenter n°2 (female)		0.018	-0.018
			(0.036)	(0.033)
	Experimenter n°3 (male)		0.100***	0.096***
			(0.038)	(0.029)
	First session in the day		0.048	0.070**
			(0.030)	(0.032)
	Behavior in the 1st round of the last two sessions [a]		-0.057	0.020
			(0.200)	(0.133)
	Number of sessions held before (t)		-0.015***	-0.009***
			(0.003)	(0.003)
Observations		670	670	670

Marginal Probit for (ii)-(iv). Marginal effects reported. Robust standard errors, clustered by session. * Significant at 10%; ** significant at 5%; *** significant at 1%. [a]It is calculated as the mean of previous 2 sessions' deviation from the average contribution in the neighborhood

Table C2. Controls for the decision to contribute in the group account in the first round 2007-Table C1 (Cont.)

Dependent variable: 1 if player contributed to the group account in the first round				
	Independent Variables	(ii)	(iii)	(iv)
Session composition	Percentage of participants with less than secondary complete			0.366***
				(0.119)
	Mean absolute deviation of level of education within the session			0.305**
				(0.138)
Observations		670	670	670

Marginal Probit for (ii)-(iv). Marginal effects reported. Robust standard errors, clustered by session. * Significant at 10%; ** significant at 5%; *** significant at 1%. [a]It is calculated as the mean of previous 2 sessions' deviation from the average contribution in the neighborhood

Appendix D. Independent Cross section analysis

Table D1. Role of the leaders (first round)

Dependent variable: 1 if player contributed to the group account in the first round				
Independent Variables	IX	X	XI	XII
Dummy *Ciénaga*	0.081	0.082	0.058	0.059
	(0.057)	(0.058)	(0.052)	(0.052)
Dummy 2007	0.102**	0.100*	0.115***	0.112***
	(0.051)	(0.051)	(0.034)	(0.034)
Program impact	0.298***	0.297***	0.235***	0.234***
	(0.046)	(0.046)	(0.034)	(0.034)
1 if the participant is considered a Leader	-0.057**	-0.080**	-0.060**	-0.085***
	(0.026)	(0.034)	(0.024)	(0.033)
Percentage of participants who consider the player as Leader in the session		0.344		0.382
		(0.343)		(0.380)
Basic Controls - Participant's characteristics	Yes	Yes	Yes	Yes
Session Characteristics	No	No	Yes	Yes
Session composition (information within session)	No	No	Yes	Yes
Observations	1,384	1,384	1,384	1,384

Marginal Probit for IX-XII. Marginal effects reported. Robust standard errors, clustered by session. * Significant at 10%; ** significant at 5%; *** significant at 1%.

Table D2a. **Controls for the decision to contribute in the group account 2008 and 2007. Basic controls - participant's characteristics. Table 9 and Table D1**

Dependent variable: 1 if player contributed to the group account in the first round							
Independent Variables	VI	VII	VIII	IX	X	XI	XII
1 if the player is a woman	0.118**	0.156***	0.157***	0.114*	0.115*	0.154***	0.154***
	(0.055)	(0.029)	(0.027)	(0.060)	(0.059)	(0.031)	(0.031)
If the players has a partner	0.000	0.006	0.007	-0.006	-0.006	0.003	0.004
	(0.036)	(0.033)	(0.033)	(0.036)	(0.036)	(0.033)	(0.034)
Level of education (0 to 5)	-0.001	-0.006	-0.004	0.001	0.001	-0.002	-0.001
	(0.010)	(0.009)	(0.009)	(0.011)	(0.011)	(0.010)	(0.010)
Household size	0.014**	0.012**	0.012**	0.016**	0.016**	0.013**	0.012**
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Ground floor (house)	-0.044*	-0.030	-0.029	-0.046*	-0.046*	-0.031	-0.031
	(0.025)	(0.023)	(0.023)	(0.025)	(0.025)	(0.022)	(0.022)
Number of years living in the	0.001	0.000	0.000	0.002	0.002	0.001	0.001
neighborhood	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
1 if the player's home has water	0.042	0.037	0.042	0.039	0.040	0.039	0.038
pipe access	(0.034)	(0.032)	(0.031)	(0.036)	(0.036)	(0.032)	(0.032)
Sewage	-0.041	-0.032	-0.026	-0.037	-0.038	-0.021	-0.021
	(0.029)	(0.027)	(0.027)	(0.029)	(0.029)	(0.027)	(0.027)
No electricity	0.146**	0.100	0.110	0.126*	0.125*	0.092	0.094
	(0.072)	(0.072)	(0.073)	(0.072)	(0.071)	(0.072)	(0.073)
1 if the player is the head of	0.021	0.019	0.021	0.015	0.016	0.019	0.019
household	(0.039)	(0.039)	(0.039)	(0.039)	(0.039)	(0.039)	(0.039)
Age	0.005***	0.005***	0.005***	0.005***	0.005***	0.005***	0.005***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
1 if the player is unemployed	-0.013	0.010	0.011	-0.018	-0.017	0.007	0.007
	(0.058)	(0.058)	(0.058)	(0.057)	(0.057)	(0.058)	(0.058)
1if the player has her own housing	-0.046*	-0.054**	-0.053**	-0.039	-0.039	-0.045*	-0.045*
	(0.027)	(0.025)	(0.026)	(0.027)	(0.027)	(0.026)	(0.026)
1 if player belongs to the fifth	0.040	0.033	0.038	0.048	0.046	0.042	0.042
quintile per capita income	(0.030)	(0.029)	(0.029)	(0.031)	(0.031)	(0.031)	(0.031)
1 if the player has a phone	-0.004	-0.003	-0.006	-0.000	-0.000	-0.004	-0.004
	(0.038)	(0.039)	(0.038)	(0.039)	(0.039)	(0.039)	(0.039)
Number of rooms in the house	-0.016	-0.011	-0.010	-0.017	-0.017	-0.010	-0.010
	(0.014)	(0.014)	(0.015)	(0.014)	(0.014)	(0.015)	(0.015)
1 if the player is displaced (self-	0.027	0.018	0.015	0.030	0.030	0.018	0.017
declared)	(0.032)	(0.032)	(0.032)	(0.032)	(0.032)	(0.032)	(0.033)
1 if the player declares she	-0.048	-0.041	-0.042	-0.047	-0.048	-0.044	-0.044
understood everything	(0.030)	(0.030)	(0.030)	(0.029)	(0.030)	(0.030)	(0.030)
1 if the household receives other	-0.059**	-0.053*	-0.050*	-0.061**	-0.062**	-0.049*	-0.048*
support from another institution	(0.030)	(0.027)	(0.027)	(0.029)	(0.029)	(0.027)	(0.027)
Observations	1384	1384	1384	1,384	1,384	1,384	1,384

Marginal Probit effects reported. Robust standard errors, clustered by session. * Significant at 10%; *** significant at 5%; *** significant at 1%.

Table D2b. **Controls for the decision to contribute in the group account 2008 and 2007. Session characteristics**

Dependent variable: 1 if player contributed to the group account in the first round				
Independent Variables	VII	VIII	XI	XII
Number of players in session	-0.032***	-0.039***	-0.038***	-0.036***
	(0.009)	(0.008)	(0.008)	(0.008)
1 if there is at least one Man in the session	0.176***	0.207***	0.200***	0.200***
	(0.038)	(0.037)	(0.046)	(0.045)
Experimenter n°2 (female) in 2007	-0.045	-0.071	-0.077*	-0.073*
	(0.052)	(0.044)	(0.040)	(0.042)
Experimenter n°3 (male) in 2007	0.116**	0.130***	0.134***	0.134***
	(0.046)	(0.041)	(0.043)	(0.044)
Experimenter n°2 (female) in 2008	0.160***	0.138***	0.101**	0.088
	(0.046)	(0.040)	(0.050)	(0.054)
First session in the day	0.081**	0.095***	0.100***	0.096***
	(0.037)	(0.032)	(0.036)	(0.035)
Behavior in the 1st round of the last two sessions [a]	-0.224	-0.142	-0.210*	-0.233*
	(0.174)	(0.143)	(0.123)	(0.125)
Number of sessions held before (t)	-0.018***	-0.016***	-0.018***	-0.019***
	(0.002)	(0.002)	(0.003)	(0.003)
Observations	1384	1384	1,384	1,384

Marginal Probit effects reported. Robust standard errors, clustered by session. * Significant at 10%; *** significant at 5%; *** significant at 1%. [a]It is calculated as the Mean of previous 2 Sessions' Deviation from the average contribution in the neighborhood

Table D2c. **Controls for the decision to contribute in the group account 2008 and 2007. Session composition**

Dependent variable: 1 if player contributed to the group account in the first round			
Independent Variables	VIII	XI	XII
Percentage of participants with less than high school diploma	0.366***	0.387***	0.351***
	(0.090)	(0.110)	(0.126)
Mean absolute deviation of level of education in the session	0.176*	0.195*	0.176
	(0.099)	(0.105)	(0.111)
Observations	1384	1,384	1,384

Marginal Probit effects reported. Robust standard errors, clustered by session. * Significant at 10%; *** significant at 5%; *** significant at 1%

Appendix E. Descriptive Statistics

Variable		Base line					Follow-up (Independent Cross Section)				
		mean	**s.d***	**max**	**min**	**N**	**mean**	**s.d***	**max**	**min**	**N**
Basic Controls - Participant's characteristics											
Decision to contribute		0.20	0.40	0	1	676	0.29	0.46	0	1	714
1 if the player is a woman		0.99	0.10	0	1	676	0.98	0.12	0	1	714
If the players has a couple		0.70	0.46	0	1	676	0.72	0.45	0	1	714
Level of Education (0 to 5)		2.78	1.27	0	5	676	2.75	1.31	0	5	714
Household size		5.72	2.11	2	15	676	5.67	2.02	2	15	714
Ground Floor (house)		0.33	0.47	0	1	676	0.28	0.45	0	1	714
Number of Years living in the neighborhood		18.38	9.42	1	50	676	18.55	9.78	0	57	714
1 if the player's home has water pipe access		0.86	0.35	0	1	676	0.88	0.33	0	1	714
Sewage		0.39	0.49	0	1	676	0.46	0.50	0	1	714
No Electricity		0.04	0.19	0	1	676	0.03	0.18	0	1	714
1 if the player is the Head of Household		0.26	0.44	0	1	676	0.33	0.47	0	1	714
Age		35.83	9.94	18	75	674	36.23	9.75	16	74	714
1 if the player is Unemployed		0.07	0.25	0	1	676	0.04	0.20	0	1	714
1if the player has her Own Housing		0.71	0.45	0	1	676	0.59	0.49	0	1	714
1 if player belongs to the fifth quintile per capita income		0.20	0.40	0	1	676	0.18	0.38	0	1	714
1 if the player has a Phone		0.18	0.38	0	1	676	0.18	0.38	0	1	714
Number of rooms in the house		2.28	1.05	1	6	675	2.29	1.07	0	6	714
1 if the player is Displaced (self-declared)		0.14	0.34	0	1	675	0.13	0.34	0	1	714
How much the player is Player understood the instructions		0.73	0.44	0	1	675	0.67	0.47	0	1	714
1 if the household receives support from another institution		0.13	0.34	0	1	675	0.10	0.30	0	1	714
Session Characteristics											
Number of players in session		24.26	1.43	21	28	676	24.65	0.78	22	25	714
1 if there is at least one Man in the session		0.14	0.35	0	1	676	0.31	0.46	0	1	714
Experimenter n°2 (female) in 2007		0.17	0.38	0	1	676	0.00	0.00	0	0	714
Experimenter n°3 (male) in 2007		0.33	0.47	0	1	676	0.00	0.00	0	0	714
Experimenter n°2 (female) in 2008		0.00	0.00	0	0	676	0.62	0.49	0	1	714
First Session in the Day		0.15	0.36	0	1	676	0.34	0.48	0	1	714
Behavior in the 1st round of the last two sessions		0.01	0.07	-0.17	0.18	676	0.08	0.13	-0.11	0.45	714
Number of sessions held before (t)		7.53	4.02	1	14	676	8.89	6.25	1	24	714
Percentage of participants with less than secondary complete		0.60	0.13	0.32	0.84	676	0.71	0.09	0.52	0.92	714
|Std Dev of level of education in in the session - mean of level of education in the neighborhood in that year|		1.53	0.16	1.26	1.87	676	1.47	0.14	1.23	1.71	714
Social Capital measures											
Membership in at least one organization		0.39	0.49	0	1	676	0.27	0.44	0	1	714
Participation in neighborhood decisions		0.54	0.50	0	1	676	0.41	0.49	0	1	714
Participation in the neighborhood meetings		0.62	0.49	0	1	660	0.48	0.50	0	1	714
Vote Local Elections (2003 and 2007)		0.52	0.50	0	1	676	0.72	0.45	0	1	714
Vote Presidential Elections (2006)		0.67	0.47	0	1	676	0.78	0.42	0	1	714
Trust	Most people	0.07	0.25	0	1	642	0.07	0.25	0	1	687
	Few people	0.62	0.49	0	1	642	0.62	0.49	0	1	687
	None	0.31	0.46	0	1	642	0.31	0.46	0	1	687
Perception within the community	Cooperation	0.26	0.44	0	1	620	0.32	0.47	0	1	691
	Reciprocity	0.24	0.43	0	1	620	0.14	0.35	0	1	691
	Selfishness	0.49	0.50	0	1	620	0.54	0.50	0	1	691
Player is identified as leader in the session		0.33	0.47	0	1	676	0.18	0.39	0	1	714
Percentage of players identified as leader in the session		0.03	0.05	0	0.36	676	0.01	0.03	0	0.28	714
Average number of relatives		0.14	0.41	0	3	676	0.13	0.40	0	3	714
Average number of friends		1.04	2.21	0	17	676	1.46	1.83	0	11	714
Average number of acquaintances		7.74	7.60	0	27	676	0.44	1.11	0	10	714
Average number of connections		8.91	7.75	0	27	676	2.04	2.12	0	13	714
Relatives density a		0.01	0.01	0	0.02	676	0.01	0.01	0	0.04	714
Friendship density		0.05	0.06	0	0.30	676	0.11	0.04	0.06	0.25	714
Acquaintances density		0.37	0.14	0.18	0.66	676	0.03	0.02	0	0.08	714
Connections density		0.43	0.15	0.20	0.72	676	0.16	0.05	0.08	0.33	714
Player claims to be a ML		0.17	0.37	0	1	342	0.05	0.22	0	1	714
There is at least a ML in the session		0.51	0.50	0	1	676	0.79	0.41	0	1	714

* s.d. means standard deviation

Appendix References

Attanasio, Orazio; Luca Pellerano and Sandra Polanía-Reyes. 2009. "Building Trust? Conditional Cash Transfer Programs and Social Capital." Fiscal Studies, 30(2), pp. 139.

Cárdenas, Juan Camilo and Christian Jaramillo. 2007. "Cooperation in Large Networks: An Experimental," In Documentos CEDE, Universidad de los Andes-Cede. Bogota.

CCV, 2011. Summary of quality of life in Locality 2 in Cartagena. [Cartagena como vamos]. http://www.cartagenacomovamos.org/temp_downloads/Resumen%20Calidad%20de%20Vida%20Localidad%20II.pdf

MESEP, 2012. Mission for Splicing Series Employment, Poverty and Inequality (MESEP). Monetary poverty in Colombia: New methodology and figures 2002-2010 [Misión para el Empalme de las Series de Empleo, Pobreza y Desigualdad (mesep). Pobreza monetaria en Colombia: Nueva metodología y cifras 2002-2010] Report. DNP, DANE. Bogota, Colombia.

Molyneux, Maxine, 2011, Cash transfers, gender equity and women's empowerment in Peru, Ecuador and Bolivia, Gender & Development Vol. 19, Iss. 2, 2011

Pérez Gerson and Irene Salazar 2007. Poverty in Cartagena: an analysis by neighborhoods. [La pobreza en Cartagena: Un análisis por barrios]. Working papers on regional Economics. No. 94. Central Bank of Colombia and CEER August.

Concluding remarks

The past two decades have witnessed an explosion of new work in experimental and behavioral economics exploring and sometimes taking exception to the psychological underpinnings of the assumption that individuals are self-regarding. Even when regarding individuals as maximizers of an expected utility function, this function need not depend on individual consumption alone. Many then raise the "so what?" question: do the experimental findings really change how economists should address the traditional questions of interest in such fields as labor economics, finance, development economics? We address this question with particular reference to public economics and mechanism design and conclude that incentives may alter social preferences in forms that claim further rigorous exploration.

First, incentives sometimes crowd out noneconomic motives. This may degrade prosocial behavior and negatively affect economic performance. The present study proposes and explores the idea that what accounts for crowding out is the meaning of the fines or subsidies to the target of the incentives; this depends on the social relationships among the actors, the information the incentive provides and the preexisting normative frameworks of the actors.

Incentives that activate the target's desire to constitute himself or herself as a dignified and autonomous individual who is treated fairly by others are more effective in practice than they would be under separability between motives and incentives (see the case of reciprocators among CPR users in Colombia). The same incentives deployed by individuals who do not stand to benefit personally, and that are intended to foster pro-social behavior, such as noneconomic incentives used in the CPR game and cheap-talk communication used by the leader in the public goods game, are more likely to be complements for social preferences, crowding them in rather than out. They do this by activating rather than diminishing the target's constitutive motives such as the desire to be treated fairly and to treat others fairly, to be a good member of a community, and the feeling of shame when others regard one as having failed in doing so.

As we propose in chapter one, a policy in which incentives play a part should let the target understand that the desired modification in her actions will serve to implement an outcome that is socially beneficial so that that the target is more likely to endorse the purpose of the incentive, rather than being offended by it as either unjust or a threat to her autonomy or in some other way reflecting badly the intentions of the planner. This may also explain the success of the subsidy among the CPR villagers.

Second, we study how the existence of different social preferences within a community can be used to overcome social dilemmas and the performance of economic incentives in terms of achieving social efficiency. We explore two sources. On one hand, people may differ both in their socio-economic characteristics, their perceptions and beliefs on the social dilemma and social capital within the community, which are exogenous for the policy designer. On the other hand, the composition of social preferences in the surrounding community, which is revealed in the day-by-day interaction, affects not only the social norms but also beliefs about others' preferences on the social dilemma.

Finally, further research on the formation on social preferences should shed light on the underlying social and psychological mechanisms, which are beyond the scope of the present

study. Identification, which is the key issue in the study of social preferences, has therefore a time component to it as well as the (previously discussed) cross-sectional component. Addressing identification calls for stronger theoretical underpinnings for field experiments as well as more sophisticated econometric techniques to understand the data that is gathered from them.

References

Datta, S., & Mullainathan, S. 2014. Behavioral Design: A New Approach to Development Policy. Review of Income and Wealth, 60(1), 7-35.