



**XITH URRUTIA ELEJALDE SUMMER SCHOOL
ON ECONOMICS AND PHILOSOPHY**

Social Norms

Abstracts: Contributed papers + posters

Contributed papers

Monday 14

Giacomo Sillari (UPenn), "Conventions, Norms and Rule following"

Brian Gunia (Northwestern U.), "Acting out what's right: the creation and development of intragroup moral norms"

Angela Milano de Oliveira, Rachel Croson & Catherine Eckel (U. of Texas at Dallas), "Understanding Pro-social behavior in the field: experimental evidence from a low income neighbourhood"

Tuesday 15

Rebekka A. Klein (URPP Foundations of Human Social behavior), "Altruistic Punishment and Norm enforcement"

Stefania Ottone (EconomEtica, U. Eastern Piedmont), Ferruccio Ponzano (U. Eastern Piedmont), Luca Zarri (U. Verona), "Third-Party Punishment, Metanorms and Subjective Fairness"

Christian Traxler (Max Planck Institute), Joachim Winter (U. Munich), "Survey Evidence on Conditional Norm Enforcement"

Wednesday 16

Erte Xiao and Cristina Bicchieri (U. Pennsylvania), "When Equality Trumps Reciprocity: Evidence from a Laboratory Experiment"

Antonella Marchetti, Ilaria Castelli (U. Cattolica del Sacro Cuore), Alan G. Sanfey (U. Arizona), "Intentionality and Decision-Making in Children: A Research with the Ultimatum Game"

Marco Faillo (U.Trento), Stefania Ottone (EconomEtica and U. Eastern Piedmont), Lorenzo Sacconi (U.Trento and EconomEtica), "Compliance by believing: an experimental exploration on social norms and impartial agreements"

Erin Krupka, Roberto Weber and Rachel Croson, ""When in Rome": Applying a method for eliciting social norms"

Poster presentations

Monday 14

Julia Cordero Coma (Juan March Institute), "The Role of Interpersonal Communication in the Formation of Social Norms Regulating Preventive Sexual Behaviour against HIV/AIDS"

Azi Lev-On (Ariel University Center), "Communication, Trust and Reciprocity among Members of Ethnic Groups in Conflict"

Donna Harris (U. Cambridge), "Social Connections, Group behaviours and Corruption: An Experimental Study of a New 'Favour Game'"

Brice Corgnet (U. Navarra), Angela Sutan (Burgundy Business School), Robert Veszteg (Universidad del País Vasco), "Team Formation Experiments and the Fairness Norm: An International Comparison"

Tuesday 15

Ryan Muldoon (U. Pennsylvania), "Agreeing to Disagree: How Conflicting Norms Can be Mutually Reinforcing"

Alessandra Smerilli (U. East Anglia), "The emergence of cooperation in a heterogeneous world. An evolutionary approach"

Benoît Dubreuil (FNRS, U. libre de Bruxelles), "The logic of collective sanction"

Giuseppe Danese, Luigi Mittone (U. Trento), "Reciprocity, Exchange and Redistribution. An experimental investigation inspired by Karl Polanyi's 'The Economy as Instituted Process'"

Conventions, Norms, and Rule-Following

Giacomo Sillari

¹ 313 Logan Hall
PPE Program
University of Pennsylvania
gsillari@sas.upenn.edu

In a long-standing philosophical lineage that stretches back to David Hume, conventions are a central element of philosophical analysis. In a conventionalist approach, for instance, justice consists of a system of actions that are performed by members of society because general conformity to the system is individually advantageous for them. The notion of convention largely informs the philosophical discussion of topics ranging from morality to linguistic meaning. In the social sciences, institutions as property or money are also thought of as conventions. An important problem for this view concerns the genesis of conventions, since there are in general multiple ways for the agents to coordinate their actions. Consider the example ([Lew69]) of a convention gradually emerging in a community such that, if a telephone conversation is cut off, the original caller calls back while the other person waits. For a stable mutually beneficial outcome to arise, agents need to recognize similarities in successive instances of the coordination problem, and need to project past regularities of the behavior into the current coordination problem. These, however, are cognitive capacities that the purely game-theoretic accounts currently used to model conventions fail to consider.

Among scholars, positions vary rather markedly when it comes to explicate the mechanisms through which conventions are selected and maintained. To address the problem of selecting a convention, [Lew69] stresses the epistemic role of precedent and common knowledge (cf. [VS05], [CS03]) while Skyrms ([Sky96], [?]) and Binmore ([Bin94], [Bin98], [Bin05]) address the problem through evolutionary analysis. Both Sugden ([?]) and Bicchieri ([Bic06]), although with different emphasis, point at the importance of framing, salience and other cognitive factors, without however explicitly modeling them in their accounts of norms and conventions.

While several strands of literature have attacked the problem of choosing a convention from different perspectives, the various attempts have reached unsatisfactory conclusions by ignoring essential elements of the overall picture. In philosophy, some scholars claim that social conventions must be explained in terms of collective notions that cannot in turn be accounted

for by appealing to the strategic behavior of isolated interacting individuals. In particular, the rational actor model is insufficient to solve the selection problem and the critics claim has to be abandoned altogether. In economics, evolutionary solutions based on the replicator dynamics are criticized as they are based on interacting agents entirely bereft of rationality. Yet, a full-fledged answer to the question of how we choose among alternative conventions is of great practical relevance since many possible conventions are socially harmful. Indeed, human behavior often coordinates on harmful social conventions (e.g. segregation, pollution, female genital mutilation) and policy makers often face the problem of steering members of society towards beneficial social behavior. The reason why philosophers, economists and social scientists alike have failed to provide a satisfactory solution depends on the nature of coordinating behavior, which cannot be analyzed in isolated parts individually studied by different disciplines. The problem of convention selection is akin to that of belief formation: since each agent has an incentive to conform, she will do whatever she believes others will do. But then classical game theory and epistemic analysis are by themselves insufficient to provide an answer, since neither are concerned with belief formation; so is evolutionary analysis, since it is not concerned with rational agents. On the other hand, cognitive psychology, if not game-theoretically informed, fails to account for the difference between mere cognitive regularities in belief formation and effective social conventions.

There is an important link between social norms and social conventions, largely revolving around the question whether the expectations underlying conventional behavior are normative. To understand and elucidate such a link, I will address a broader topic in analytic philosophy, namely Wittgenstein's rule-following considerations. I intend to show that rule-following in the sense of Wittgenstein can be understood in the game-theoretic, Lewisian framework. In so doing, I will argue that the solution to the skeptical paradox triggered by rule-following finds an analogous counterpart in the solution to the equilibrium selection problem triggered by the game-theoretic account of convention. Besides its general philosophical significance the analysis will provide a firmer justification for my claim that social conventions in the sense of Lewis contain elements extraneous to individual agency.

References

- [Bic06] Cristina Bicchieri. *The Grammar of Society*. Cambridge University Press, Cambridge, 2006.
- [Bin94] Ken Binmore. *Game Theory and the Social Contract. Playing Fair*,

volume 1. MIT Press, 1994.

- [Bin98] Ken Binmore. *Game Theory and the Social Contract. Just Playing*. MIT Press, 1998.
- [Bin05] Ken Binmore. *Natural Justice*. Oxford University Press, 2005.
- [CS03] Robin P. Cubitt and Robert Sugden. Common knowledge, salience and convention: a reconstruction of David Lewis' game theory. *Economics and Philosophy*, 19:175–210, 2003.
- [Lew69] David Lewis. *Convention: A Philosophical Study*. Harvard University Press, Cambridge, Mass., 1969.
- [Sky96] Brian Skyrms. *The Evolution of the Social Contract*. Cambridge University Press, Cambridge, Mass., 1996.
- [VS05] Peter Vanderschraaf and Giacomo Sillari. Common knowledge. *Stanford Encyclopedia of Philosophy*, 2005.

**ACTING OUT WHAT'S RIGHT:
THE CREATION AND DEVELOPMENT OF INTRAGROUP MORAL NORMS**

Brian Gunia

Doctoral Student
Kellogg School of Management
Northwestern University

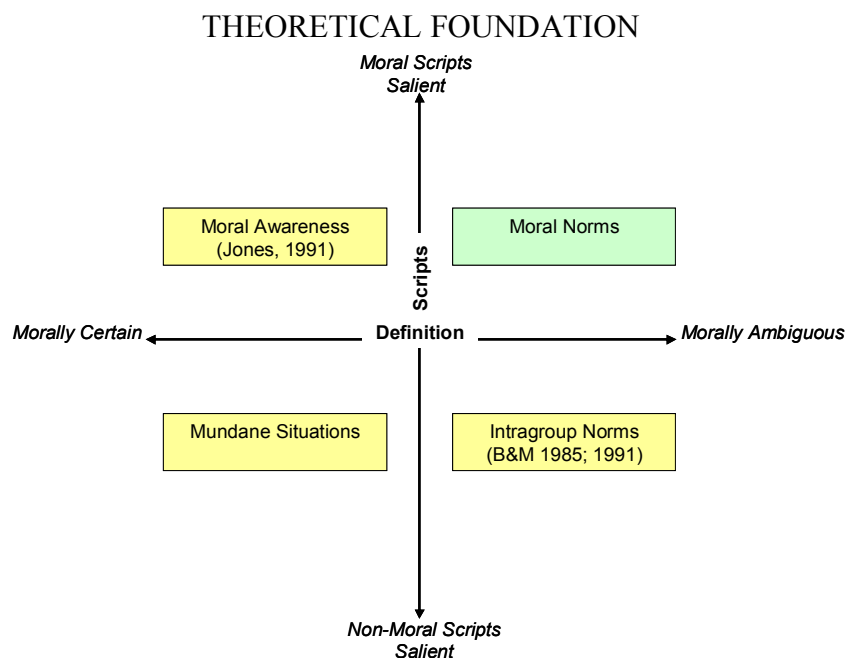
Submitted December 2007
XI Summer School on Economics and Philosophy

ACTING OUT WHAT'S RIGHT: THE CREATION AND DEVELOPMENT OF INTRAGROUP MORAL NORMS

OVERVIEW

Research on social norms often concludes that norms influence behavior by translating society's moral judgments into specific social guidelines (e.g., Cialdini, Kallgren, & Reno, 1991; Cialdini, Reno, & Kallgren, 1990; Gouldner, 1960). Although undeniably important, these top-down processes divert attention from a bottom-up possibility – that social behavior could actually *produce* norms dictating moral right and wrong, which eventually influence the broader society.

In the current proposal, three studies examine whether and when social interactions produce norms dictating right and wrong. We suggest that, when people define a situation as morally ambiguous, they seek social cues for guidance on how to behave. When others' behavior stimulates moral scripts, a *moral* norm can develop. This norm identifies the behaviors understood to be right, in a moral sense. The development of a moral norm involves the same kinds of interpersonal dynamics as the development of a standard, intragroup norm (Bettenhausen & Murnighan, 1985, 1991). When new situations or interactions challenge an established moral norm, however, we expect people to protect and retain the moral norm, in its original form, more than they would an intragroup norm. The reason is that the scripts underlying moral norms (unlike those underlying intragroup norms) make the person's ought or ideal self (Higgins, 1987) salient. This should increase their resistance to a new norm that would create an aversive self-discrepancy. If people do try to protect their identities by maintaining a norm, identification and motivation processes might then take a more important place in models of norm formation.



This chart depicts four ways that people can interpret the moral features of a new situation. Following Bettenhausen & Murnighan (1985), the chart distinguishes between definitions of a current situation (along the X-axis) and behavioral scripts informed by past situations that form the basis of behavior in the situation (along the Y-axis). Definitions of a current situation range from

moral certainty (in which individuals have almost no doubts about the right course of action) to moral ambiguity (in which they have almost no clue). Bettenhausen & Murnighan (1985) also suggested that situations and the behavior of people within them can render a variety of scripts – ranging from moral to non-moral – salient. We distinguish moral norms from intragroup norms by the situational definitions and scripts that they stimulate in the people involved. In short, we suggest that moral norms develop in morally ambiguous situations when the behavior of others makes moral scripts salient. Once morality surfaces as an issue, individual identification processes make the resulting norms relatively impervious to change.

EXPERIMENT 1: NORMS THROUGH TIME

In this experiment, we use a confederate to create a moral norm. We then investigate whether this norm constrains subsequent individual behavior more than an intragroup norm.

Procedure

Experiment 1 uses a prisoner's dilemma game in which participants interact with a confederate via instant messaging (IM) technology. The confederate makes pre-tested statements to establish a moral or intragroup, cooperative or competitive, norm. Participants then make an independent recommendation to a future player on how to play the game, and they play it independently for six more rounds.

Prior to the experiment, participants are randomly assigned to one of four conditions:

		<u>Type of argument</u>	
		Moral	Non-moral
<u>Content of argument</u>	Cooperative	1	2
	Competitive	3	4

Each half-hour, five participants and one confederate arrive at the lab and sign consent forms. The experimenter informs them that the experiment is concerned with online communication and individual decision making. The experiment, they learn, involves a “practice session” during which two participants will be randomly paired. The pair will face an ambiguous situation in which they work together, exchanging instant messages to reach the best decision. Unbeknownst to participants, the experimenter pairs each one with the confederate. Following practice, participants will make a series of “real decisions” on their own. To increase realism, the experimenter concludes these instructions by taking each participant's picture so that everyone, while interacting, can see a picture of their partner.

Next, the experimenter leads each participant into a private breakout room. Participants watch while the experimenter “randomly” pairs them with the confederate (whose picture appears on the screen) and launches an instant message (IM) session with him/her. The experimenter instructs participants to carefully read the paper packet located on their table and send an introductory IM to their partner (the confederate). Participants learn that their partner will do the same.

Instructions in the packet describe the “Truffle Purchasing Game” (Bettenhausen & Murnighan, 1991), which includes the following payoffs that result in a mildly cooperative equilibrium:

		Opponent	
		COOP	COMP
Participant	COOP	(52, 52)	(17, 85)
	COMP	(85, 17)	(41, 41)

Participants believe that they and the confederate will play an unknown number of practice rounds against another live individual or dyad. In actuality, all dyads play four rounds, and no such opponents exist. The instructions ask each dyad to reach a collective decision in each round, by sending an IM with their decision to the experimenter.

Participants always finish reading before confederates. They send an IM to the confederate, who responds with the same basic statements across conditions. The statements use moral or non-moral arguments to advocate cooperation or cooperation. These statements were extensively pre-tested to differ in moral content and recommended competitive / cooperative behavior, but not in strength. The following chart lists the confederates' IM statements; manipulations are italicized.

	Moral Conditions	Non-Moral Conditions
1	ok done, so heres what I think...	ok done, so heres what I think...
	<i>To me this is a moral situation...</i>	<i>This seems like a straightforward situation...</i>
	so i say we bid [low/high]. What do u think?	so i say we bid [low/high]. What do u think?
2	hmm, well... <i>I gotta say, I see this as a moral situation. Bidding [low/high] is the right thing to do...</i>	hmm, well... <i>Bidding [low/high] wouldn't make sense...</i>
	So i say we make a [low/high] bid this time.	So i say we make a [low/high] bid this time.
3	Ok. Not sure about u, but I still think we should bid [low/high]	Ok. Not sure about u, but I still think we should bid [low/high]
4	<i>Let's keep it up. Bidding [low/high] just wouldn't be right.</i>	<i>I say place a [low/high] bid again.</i>

To increase realism, confederates can also add words like, “yes,” “no,” and “not sure” or respond to clarification questions or protests using a set of pre-formulated answers. These responses are the same in all four conditions. Overall, we expect that the confederate's statements to cue moral vs. non-moral scripts in participants.

In each round, participants and confederates make their decision (high bid/competitive or low bid/cooperative) by sending an IM to the experimenter. Each time, the experimenter sends feedback indicating that the competitor team played “tit-for-tat” (Axelrod & Hamilton, 1981). After four rounds, the computer ends the practice session (because bad weather has ended the truffle growing season). The goal of these four “practice sessions” is to successfully establish a moral or intragroup, competitive or cooperative norm. We consider the relevant norm established if the confederate makes all of the pre-tested statements and the dyad's last two choices match the confederate's recommendation.

Following the practice sessions, participants make a series of individual decisions on two (counterbalanced) tasks. First, they learn that a future participant will play the same game but

without any practice sessions. Rather, they will only receive the current participant's guidance on what to do. Instructions ask current participants to indicate "how you think the game should be played" and explain their rationale. After participants complete this task, they play an unknown number of additional rounds of the same game, individually, against an unknown counterpart. In reality, the computerized opponent always makes tit-for-tat choices; the experiment ends after six rounds. Participants then complete a post hoc questionnaire. Finally, we probe for suspicion, debrief participants in-full, and pay them \$10 for their time.

In both the recommended strategy and individual play tasks, we predict that more moral norm participants will continue to espouse their norm, relative to intragroup norm participants.

EXPERIMENT 2: STRUCTURAL CHALLENGE

This experiment investigates the possibility that moral and intragroup norms will differentially constrain behavior, despite structural changes that encourage counter-normative behavior.

Procedure

Participants in this experiment follow the same procedure as in experiment 1. The only difference is that, during the individual task, participants play a different game than the original. Those in the cooperative conditions (1 and 2) face a strong, competitive structural challenge, meaning that the payoffs strongly encourage competitive behavior. Those in the competitive conditions (3 and 4) face a strong, cooperative structural challenge, meaning that the payoffs strongly encourage cooperative behavior. To provide a baseline, some participants in all conditions face no structural challenge, (i.e., the payoffs remain the same as in the original game). The following figure outlines the structural challenges participants may encounter:

Structural Challenge

	Strong Comp	Neutral	Strong Coop
1	1-comp	1-neutral	
2	2-comp	2-neutral	
3		3-neutral	3-coop
4		4-neutral	4-coop

Payoffs for each of the conditions in this experiment – where higher g-values indicate more competitive situations (Murnighan & Roth, 1983) – follow:

Payoffs for Strong Competitive Challenges (1-comp and 2-comp); $g = -15$

		<u>Opponent</u>	
		COOP	COMP
<u>Participant</u>	COOP	(40, 40)	(8, 70)
	COMP	(70, 8)	(37, 37)

Payoffs for Neutral Challenges (1-neutral through 4-neutral); $g = 22$ (same as original payoffs)

		<u>Opponent</u>	
		COOP	COMP
<u>Participant</u>	COOP	(52, 52)	(17, 85)
	COMP	(85, 17)	(41, 41)

Payoffs for Strong Cooperative Challenges (3-coop and 4-coop); $g = 58$

		<u>Opponent</u>	
		COOP	COMP
<u>Participant</u>	COOP	(48, 48)	(8, 70)
	COMP	(70, 8)	(32, 32)

Participants again play six individual rounds with their new payoff structure. The computerized opponent again plays tit-for-tat. Following six rounds of the game, participants answer post hoc questions and receive a suspicion check, debriefing, and compensation. We expect that the moral norm participants will continue to espouse their norm more than the intragroup norm participants, regardless of the structural challenge they face. In addition, the intragroup norm participants provide an opportunity to replicate Bettenhausen & Murnighan (1991), who found that participants who faced a competitive structural challenge (but not those who faced a cooperative one) forsook their norm.

EXPERIMENT 3: INTERPERSONAL CHALLENGE

In a third experiment, we investigate whether moral and intragroup norms will continue to constrain behavior in the face of interpersonal interaction with someone who disagrees. In this study, we observe whether an intragroup or moral norm survives when two people advocating norms that differ in moral and/or behavioral content interact.

Procedure

Participants in this experiment initially follow the same procedure as in experiment 1. Following the practice sessions, however, the experimenter leads two participants at a time to new breakout rooms. The recommendation and individual play procedures from experiment 1 are replaced by interactions with another real, live participant. Dyads are created according to the participants' original condition numbers. The following charts reiterate the original condition numbers and illustrate how participants from the various conditions are paired:

Original Conditions

		Type of argument	
		Moral	Non-moral
Content of argument	Cooperative	1	2
	Competitive	3	4

Experiment 3 Pairings

1	3
2	4
1	4
2	3

The first and second pairings combine participants who were both exposed to a moral or to an intragroup norm. In each case, however, the norm advocates a different behavioral strategy (cooperation vs. competition). The third and fourth pairings combine participants exposed to norms that differ in both moral content and behavioral strategy.

Participants then play the same Truffle Purchasing Game that they did during the “practice sessions.” This time, the instructions describe the game as “real” and introduce no structural challenges. Participants are not told how many rounds they will play, which in this case is ten. Participants then respond individually to the post hoc questions and receive a suspicion check, debriefing, and compensation. We predict that that 1/4 pairs will tend to choose cooperatively, and 2/3 pairs will tend to choose competitively. In line with Bettenhausen & Murnighan (1991), we also predict that 1/3 and 2/4 pairs will tend to choose cooperatively but that – because the 1/3 pairs view the interaction as moral – they will have more difficulty reaching agreement on their choices than the 2/4 pairs.

REFERENCES

- Axelrod, R., & Hamilton, W. D. (1981). The evolution of cooperation. *Science*, *211*(4489), 1390-1396.
- Bettenhausen, K., & Murnighan, J. K. (1985). The emergence of norms in competitive decision-making groups. *Administrative Science Quarterly*, *30*(3), 350-372.
- Bettenhausen, K., & Murnighan, J. K. (1991). The development of an intragroup norm and the effects of interpersonal and structural challenges. *Administrative Science Quarterly*, *36*(1), 20-35.
- Cialdini, R. B., Kallgren, C. A., & Reno, R. R. (1991). A focus theory of normative conduct: A theoretical refinement and reevaluation of the role of norms in human behavior. *Advances in Experimental Social Psychology*, *24*, 201-234.
- Cialdini, R. B., Reno, R. R., & Kallgren, C. A. (1990). A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places. *Journal of Personality and Social Psychology*, *58*(6), 1015-1026.
- Gouldner, A. W. (1960). The norm of reciprocity: A preliminary statement. *American Sociological Review*, *25*(2), 161-178.
- Higgins, E. T. (1987). Self-discrepancy: A theory relating self and affect. *Psychological Review*, *94*(3), 319-340.
- Murnighan, J. K., & Roth, A. E. (1983). Expecting continued play in prisoners dilemma games: A test of several models. *Journal of Conflict Resolution*, *27*(2), 279-300.

Understanding Pro-social Behavior in the Field: Experimental Evidence from a Low-Income Neighborhood

Angela Milano*, Rachel Croson & Catherine Eckel
University of Texas at Dallas

Abstract Submission for the XI Summer School on Economics and Philosophy: Social Norms
San Sebastian, Spain

The voluntary provision of public goods is a social dilemma, and many have argued that the characteristics that influence neighborhood quality—including safety, parks, school quality and participation in Parent Teacher Associations—rely on the effectiveness of neighborhoods at solving this dilemma. In order to understand the circumstances under which individuals are willing to cooperate with others in their neighborhood to overcome this dilemma, our research analyzes the impact of an individual's beliefs about others' social preferences on pro-social behavior. We use participants from a low-income, minority neighborhood as participants in the experiment. Interestingly, we find that an individual's beliefs about other's social preferences have a non-monotonic effect on their own behavior.

Previous experimental research has largely focused on observed levels of cooperative behavior, and attempts to explain why behavior deviates from equilibrium. The samples chosen were traditionally the convenience sample of university undergraduates (see e.g. the papers cited in Ledyard, 1995), an appropriate sample for testing theory. However, to ensure that these results were not being driven by the sample, cross-cultural research investigated the impact of sample selection on behavior (see e.g. the papers in Henrich, *et al*, 2004). Preliminary evidence suggests that experimental measures of cooperation vary with community characteristics, which opens up the potential to use these experiments as a measurement tool; as a way to measure what was previously unobserved heterogeneity at the individual and community levels. We believe that this individual- and community-level heterogeneity in the willingness to engage in pro-social behavior has the potential to explain why some communities fail to develop economically while others (that appear similar by census characteristics) manage to flourish. Our experiments are designed to measure this heterogeneity and to explore the determinants of the willingness to engage in pro-social behavior with their neighbors.

We have chosen to focus on voluntary contributions to local public goods as the expression of social preference. Our research intersects several distinct lines of research in both the experimental literature and non-experimental studies of philanthropic activities, particularly the external validity of experimental measures of preferences and the role that beliefs/perceptions play in economic decision making. We are especially interested in the impact that beliefs have on the willingness to express pro-social preferences in this domain. Two similar motivations exist for incorporating beliefs into the analysis. One of the leading explanations for cooperative behavior is that people have a preference for reciprocity (see e.g. Coates and Neilson, 2005), a social preference where beliefs are of utmost importance. Unlike altruism (where you always donate) or spite (where you never donate), beliefs play a central role in reciprocity. A similar explanation is that a large portion of the population is of the conditional cooperators type (see e.g. Fischbacher and Gächter, 2006). Conditional Cooperators give more to public goods when they believe that others are going to give as well.

When understanding the decision to contribute to public goods, we need a better understanding of the role that beliefs about other peoples' social-preference types play in the decision to cooperate. In traditional lab studies, beliefs are generally measured by asking subjects how much they think other people are contributing to the group account (see e.g. Croson, 2007). However, when individuals make decisions in a social context, other types of beliefs and/or perceptions may come into account. Rather than only exploring the role that beliefs about contributions play in the decision process, we collect information on how trustworthy, helpful and fair others in their neighborhood are believed to be. In addition, beliefs about others' donations are elicited after all decisions are made so as to avoid changes in behavior based solely on the elicitation procedure. We find that the relationship between these various types of beliefs and the expression of pro-social preferences is more complex than one might initially expect. Surprisingly, we show that believing that others are 'nicer' does not uniformly foster more cooperation.

Our research presents a unique field experiment conducted in a culturally and ethnically distinct low-income urban neighborhood in Dallas, Texas. We measure pro-social behavior using a voluntary contribution mechanism (VCM). Additionally, we measure to the willingness of individuals to contribute to neighborhood organizations through a series of donations experiments—providing the subjects with the opportunity to donate some or all of an endowment to a local charity. A further measure of pro-social behavior is self-reported individual contributions of time and/or money to charitable organizations in their everyday lives. As independent measures we elicit individuals' beliefs about their neighbors' social preferences (trustworthiness, helpfulness, and fairness). We will now take a moment to describe the design and implementation of the study in more detail.

Experimental sessions were run in June, 2007 in the Fair Park neighborhood in South Dallas, Texas. Our results are based on 190 participants who were recruited via flyers at their homes and in local stores. Participants worked through a number of incentivized tasks: the Eckel-Grossman Risk task (Eckel and Grossman, 2002, 2007); a time-preference elicitation based on Eckel, Johnson and Montmarquette (2005); a laboratory public goods game (VCM); and three versions of a donation game which were developed for this study. The tasks are followed by a social networks survey as well as a comprehensive individual survey. Each of the decision forms was explicitly designed for use with a low-literacy population, with the games presented in pictorial form with minimal text. At the end of the session, one of the tasks was randomly chosen for actual payment, as was fully explained to the participants. This paper focuses on the results from and relationship between the linear VCM and the donation experiments.

In the VCM, participants were randomly assigned into anonymous groups of three and given an endowment of \$60 which they can allocate to either their individual or a group account. In order to simplify the game, participants were given four, discrete options. They could choose to: (1) keep all \$60, (2) keep \$40 and donate \$20, (3) keep \$20 and donate \$40, or (4) donate all \$60. Since clarity was of utmost importance for this subject pool, in the experiment we describe individuals deciding how much they wanted to "put in their wallet" and how much they wanted to "put in the group account", rather than using the more abstract language often used in these instructions. This was done to minimize confusion among the subjects and had the added advantage of creating parallelism between this and the donation experiment, described below. Money in the individual account is kept by the individual. Money placed in the group account is doubled, and then divided equally among all three members of the group, regardless of their decision ($MPCR = .66$).

In the donation experiments, participants are again arranged into groups of three and face the same decision with \$60. In this game, however, the money placed in the group account was not distributed among the participants but instead donated to an organization which provides public goods for the neighborhood. We had three donation tasks, one for each of: The Martin Luther King, Jr. Family Clinic (who provide health services), The Dallas Bethlehem Center (who provide educational services for children), and The Inner-City Community Development Corporation (who provide job training services).

Beliefs are collected at the end of the experimental activity booklet, but before the networks and post-experiment surveys. Individuals are asked to state how much of the endowment they think that each of the other two individuals contributed to the group account in each of the activities. This allows us to measure not only of the average belief, but also the distribution of that belief.

Experimental sessions lasted a little more than 2 hours, and participants were paid a \$20 show-up fee plus their earnings from the experiment. One task was chosen at random for payment with subjects being informed of this. Particularly, this method was chosen to avoid portfolio effects and to make the stakes, or each game, larger rather than having several games respectively each with smaller stakes. The median per capita income in this neighborhood is approximately \$11,500. Note that in the VCM, if everyone played the dominant strategy, earnings would be $\$60 + \$20 = \$80$, approximately equal to 1.75 days wages (14 hours). If everyone played the social optimum, earnings would be $\$120 + \$20 = \$140$, approximately equal to 3 days wages (25 hours). We believe that the stakes were large enough to ensure that participants thought carefully about the problem. Between the show-up fee and the earnings from the experiment, participants' average earnings were \$79 (\$108 if you include payments to the charities), with a minimum of \$20 (the show-up fee) and a maximum of \$280.

Allow us now to summarize some of the key results from this research. First, we will discuss some of the descriptive results. We then discuss the test of the external validity of laboratory measures of pro-social preferences by examining its predictive ability in explaining the decision to contribute in both the donation experiments (where participants give money to charitable organizations) and outside the lab. We find that the VCM has significant predictive ability in both decisions, highlighting the potential to use it as a measure of pro-social preferences.

Looking at individuals' expectations of cooperation, we do find that for the VCM that beliefs about contributions are right on average. However, for all of the Donations experiments, participants significantly overestimate the amount that others will contribute to the public good. For the health public good, the average belief was \$21.39 whereas the actual average Contribution was \$18.21. Similarly, for the Childcare public good, the average belief was \$21.07 while the actual average contribution was \$18.63. Finally, for the job trailing/entrepreneurship public good, we see that the average belief is \$20.32 whereas the average contribution is only \$16.32.

Central to our discussion in this paper is the role that perceptions of the neighbors' social preferences play in these decisions. We find that the perceptions that individuals hold about their neighbors' social preferences influences their decision to contribute to charitable causes in both the donation experiments and outside of the lab. However, the impact of beliefs is not exclusively driven by reciprocity. Individuals who believe that their neighbors are more helpful are *less* likely to contribute, while individuals who believe that their neighbors are fair and/or trustworthy are *more* likely to contribute. Our research thus suggests that pro-social preferences are multi-dimensional,

and that beliefs about different dimensions of others' behavior can have differential impacts on one's own contributions.

Finally, we see that the perception of individuals in the neighborhood is related to donations behavior, but in a more complex manner than one might anticipate. Believing that you have helpful neighbors appears to be a substitute for own-giving behavior, at least for the Health public good and in the pooled data: 'If my neighbors are helpful, they will provide the public good and I don't have to.' On the other hand, believing that my neighbors are fair is a complement to own-giving: 'If my neighbors are fair, I am inspired to be fair also, and to pull my weight in provision of the public good'.

Our results provide a new and much needed descriptive account of the relationship between beliefs about others' social preferences and an individual's pro-social behavior, which has been identified as a key factor for improving low-income urban neighborhoods. Our research suggests that understanding the beliefs that individuals hold about other members of the community is vital to understanding predicting an individual's willingness to engage in pro-social behavior.

JEL Classification Codes: H41, C93, D01, Z13

Keywords: Public Goods, Field Experiment, Social Preferences, Beliefs, External Validity

*Angela C. Milano, PhD Candidate, Department of Economics, Center for Behavioral and Experimental Economic Science, University of Texas at Dallas. 800 West Campbell Road, GR31, Richardson, Texas 75080. 972-883-4880. acm041000@utdallas.edu or milano.angela@gmail.com

References

- Coats, Jennifer and William Neilson. 2005. "Beliefs about Other-Regarding Preferences in a Sequential Public Goods Game." *Economic Inquiry*. 43.3: 614-622.
- Croson, Rachel T. A. 2007. "Theories of Commitment, Altruism and Reciprocity: Evidence from Linear Public Goods Games." *Economic Inquiry*. 45 (2): 199-216.
- Eckel, Catherine & Philip J. Grossman. 2002. "Sex Differences and Statistical Stereotyping in Attitudes Toward Financial Risk." *Evolution and Human Behavior* 23(4): 281-295.
- Eckel, Catherine & Philip J. Grossman. 2007. "Forecasting Risk Attitudes: An Experimental Study Using Actual and Forecast Gamble Choices." Forthcoming, *Journal of Economic Behavior and Organization*.
- Eckel, C. C., C. A. Johnson and C. Montmarquette. 2005. "Saving decisions of the working poor: short- and long-term horizons." *Research in Experimental Economics, Volume 10: Field Experiments in Economics*, edited by Jeff Carpenter, Glenn W. Harrison, and John A. List, (Greenwich, CT: JAI Press): p. 219-260.
- Fischbacher, Urs and Simon Gächter. 2006. "Heterogeneous Social Preferences and the Dynamics of Free Riding in Public Goods." *CeDEx Discussion Paper No. 2006-01*.
- Henrich, Joseph, Robert Boyd, Samuel Bowles, Colin Camerer, Ernst Fehr, and Herbert Gintis, eds. 2004. *Foundations of Human Sociality: Economic Experiments and Ethnographic Evidence from Fifteen Small-Scale Societies*. Oxford: Oxford University Press.
- Ledyard, J. 1995. "Public Goods: A Survey of Experimental Research." In: John Kagel and Alvin Roth, eds., *The Handbook of Experimental Economics*. Princeton, NJ: Princeton University Press.

Rebekka A. Klein
URPP Foundations of Human Social Behavior
Blümlisalpstrasse 10
CH-8006 Zurich
klein@iew.uzh.ch

ALTRUISTIC PUNISHMENT AND NORM ENFORCEMENT

A Philosophical Investigation of an Empirical Concept of Human Social Behavior

1. A biological concept of altruism

The biological understanding of altruism is based on a consideration about the economy of human nature. It says: If a person is altruistic she will incur personal (material) costs to increase the fitness or material benefit of another person. The biological understanding of altruism is not due to psychological motivation or subjective benefit expectation. It works mainly on the level of actual behavior that can be observed in labs and scanners. Therefore the concept of biological altruism is easy to apply in experimental setups that study human social behavior in a game theoretical framework.

2. The correlation of norm enforcement and altruistic punishment in humans

Several experimental studies on altruism in neuroeconomics (Fehr/Fischbacher 2003 et al.) have shown that punishment plays a key role in understanding the type of pro-sociality exhibited by humans. Altruistic behaviour in humans can be far different from direct (reciprocal) or indirect (reputation-based) altruism. It is dependent on the willingness of individuals to sanction others for their norm violation or social free-riding behavior. To understand the kind of altruism that is specific to humans the science of neuroeconomics has developed a cross-disciplinary experimental design to study a behavior that is referred to as “altruistic punishment”.

The term “altruistic punishment” refers to an act of social sanctioning which reduces the payoff of a norm violator and involves a selfless personal cost to the punisher. This cost is never likely to be recovered. The punisher is therefore an altruistic person (in a biological sense), because his behavior is taken to increase the welfare of the social system and its institutions. In the perspective of neuroeconomics altruistic punishment is among the proximate (individual) causes of human evolution which can explain why human kind has developed cooperation in more complex formations like social institutions, societies and states.

3. Questions

Can evidence for an evolutionary biological altruism in humans account for such complex social behaviors and society systems as the law and penal system? To give an example: How is the behavior of “altruistic punishment” being demarcated from the behavior that was exhibited in Abu Ghraib when the institution of penalty by law became an excuse for penalty against law? In other words: In what sense is the individual act of “altruistic punishment” and norm enforcement different from spitefulness and schadenfreude?

Third-Party Punishment, Metanorms and Subjective Fairness

Stefania Ottone (EconomEtica, University of Eastern Piedmont)

Ferruccio Ponzano (University of Eastern Piedmont)

Luca Zarri (University of Verona)

January 30, 2008

1 Introduction

In the last decades, experimental research has persuasively shown that the canonical model based on material self-interest (the so called ‘selfishness axiom’) is untenable in hundreds of experiments carried out in several countries by means of a variety of game protocols. However, a lot of disagreement still exists about the exact features of non-selfish behaviours. Our experimental data show that the canonical model fails in a variety of new ways. In particular, we both confirm previous results and obtain new findings, with regard to the nature of unselfish behaviours acting as enforcement devices of a given ‘norm of fairness’. Despite their ubiquitous presence in any human society, ‘social norms’ are still poorly understood. Experimental economics can contribute to greatly enhance our understanding of how social norms emerge and be endogenously enforced. This methodology differs from field studies on social norms, due to its potential in isolating the different forces shaping norm enforcement.

Experimental subjects have often proved to be willing to display *nonstrategic punishment*, that is forms of costly sanctioning which are not driven by (more or less sophisticated forms of) material payoff maximization¹. Hence, this mechanism represents an interesting endogenous norm enforcement device. However, behavioural economics, so far, has mainly dealt with *second-party* (nonstrategic) punishment, by focusing on the

¹ For experimental evidence on so called ‘altruistic’ punishment, see e.g. Fehr and Gächter (2000) and the survey of behavioural experiments in Gintis et al. (2003).

‘vengeful’ behaviour of subjects who had been directly hurt by other players. By contrast, it is often the case that the enforcement of a given norm of fairness depends on the (nonstrategic) action of both second and *third parties*. Therefore, in this paper we focus on such norm enforcement device and wonder whether some key findings obtained within the domain of second-party punishment also carry over to third-party punishment, by addressing the following questions: does third-party punishment occur? Is third-party punishment positively related to the ‘degree of unfairness’ of the punishees? Is third-party punishment entirely nonstrategic or is it also sensitive to the cost of punishing (as second-party punishment appears to be)?

2 The experimental design

The experimental design consists of three treatments: the Dictator Game Treatment (DGT), the Third-party Punishment Game Treatment (TPGT) and the Metanorm Treatment (MT). In the DGT the tool is the classic Dictator Game. At the beginning of the session each subject is randomly assigned a role (A or B) and groups of 2 participants are formed. In each group, participant A (the Dictator) and participant B (the Receiver) play a Dictator Game.

In the TPGT, our vehicle is the ‘third-party punishment in the dictator game’ (TP-DG) originally proposed by Fehr and Fischbacher (2004; p. 66), that is a DG in which a third player with a punishment option is introduced. At the beginning of the first stage each subject is randomly assigned a role (A, B or C) and groups of 3 participants are formed. In each group, participant A (the Dictator) and participant B (the Receiver) play a Dictator Game. In the second stage, participant C (the Observer) enters the game and has to decide whether to bear a cost in order to sanction A or to keep the whole initial endowment. This design reflects the idea that violation of a given behavioural standard may be punished not only by ‘second parties’ (participant B, in our design), but also by uninvolved ‘third parties’ (participant C).

In the MT, we study a variant of the Third-party Punishment Game, based on the notion of *metanorm* (Axelrod, 1986). After that players participate in the Third-party Punishment Game, a third stage begins. In this stage participant B has the possibility to

become an active player, by punishing participant C. The MT may be seen as a combination of the Third-party Punishment game and the well-known Ultimatum Game (UG; see Guth et al., 1982): like in the TPGT, the Observer can punish the Dictator at a cost and, like in the UG, the Receiver can punish the ‘first party’. The key difference between the MT and the UG is that in the latter the Receiver can *directly* (though implicitly, that is by rejecting the offer) punish his coplayer², whereas in the former the Receiver is only allowed to *indirectly* punish the first party by punishing the Observer for not punishing (enough) the first party.

In each treatment, A’s and C’s initial endowment is the same (20 tokens), while B’s initial endowment is 10 tokens. The cost for participant C to punish participant A for the amount of 2 tokens, is 1 token. In the MT, the cost for participant B to punish participant C for the amount of 2 tokens, is 1 token as well³. Each token’s value is 0.50 Euro.

3 The experimental procedure

The experiment was run in the Experimental Economics Laboratory (EELAB), at the University of Milano-Bicocca in Milan, Italy. The experiment was programmed and conducted with the software *z-Tree* (Fischbacher, 2007). Overall, 2 sessions for each treatment were run, with a total of 157 participants (40 participants in the DGT, 60 in the TPGT, 57 in the MT). At the beginning of the experiment, participants were informed about the sequential nature of the game protocol. The instructions were read by participants on their computer screen while an experimenter read them loudly. After reading the instructions and before subjects were invited to take decisions, some control questions were asked in order to be sure that players understood the rules of the game. At the end of each session, subjects were asked to fill in a brief survey to check for socio-demographic data. Each subject participated in one session only and partners’ identities were unknown even when the experiment was over. The strategy method at the Observer’s stage was

² The UG can then be seen as the most famous example of experimental analysis of second-party sanctions.

³ Only transfers of entire tokens are allowed and no participant can earn a negative payoff.

implemented⁴. Each session lasted for about 20 minutes for the DGT, 40 minutes for the TPGT, and about 50 minutes for the MT. Each subject earned on average 7.4 Euros.

4 Expected results

In the TPGT, we expect that selfish third parties never punish, since subjects never meet one another more than once and punishment is costly for them. For the same reason, the same expectation holds for both Observers and Receivers in the MT. In other words, our design completely rules out the possibility of both self-interested third-party sanctions and self-interested punishment on the part the Observers in the MT. Then, we also expect that, in all three treatments, if A believes that B and C are selfish, a selfish Dictator transfers nothing to the Receiver.

According to the Fehr and Schmidt (FS, 1999) model, only in the DGT transfers from player A to player B (if $\beta \geq 0.5$) is possible, while both in the TPGT and in the MT neither transfer nor punishment are predicted. We obtain the same results if we consider the models of Bolton and Ockenfels (BO, 2000) and Charness and Rabin (CR, 2002).

To sum up, according to FS, BO and CR we expect only transfer in DG, while nothing is predicted in the other two treatments.

5 Results

Result 1. *Both in the TPGT and in the MT, the punishment level on the part of the Observer decreases as the Dictator's transfer increases*

This result (see Figure 1) is perfectly in line with the experimental evidence obtained in other works (see for instance Fehr and Fischbacher, 2004⁵; Bernhard, 2005; Ottone,

⁴ When the strategy method is used, subjects are asked to state their decision in correspondence of each possible case. In our experiment, this meant that C was asked to indicate the number of deduction points for each of A's possible transfer levels before knowing A's actual choice. The final payoff was then determined on the basis of A's actual choice.

⁵ In particular, Fehr and Fischbacher (2004) find that almost two-thirds of the third parties indeed punished the violation of the distribution norm and that their punishment increased the more the norm was violated.

2005, 2007). A random effect Tobit regression of punishment on the Dictator's transfer confirms the positive relation between the level of punishment and the degree of unfairness ($p = 0.000$).

Result 2. *Metanorms do not increase the Observer's punishment levels*

When we add the possibility for the Receiver to punish the Observer (e.g. if s/he thinks s/he did not sanction enough an unfair Dictator), the Observers' behaviour does not change (see Figure 2). If we compare the Observer's level of punishment at each level of the Dictator's transfer (see Table 1), we find out that there is no significant difference (Mann-Whitney test; $p > 0.14$). A random effect Tobit regression confirms this result ($p = 0.35$).

Result 3. *Normative beliefs*

After that subjects' decisions are taken, their first-order normative beliefs are elicited. Most people think that the right transfer from the Dictator to the Receiver is different from 0 (see Table 3). If we compare the average right transfer according to the Dictators, we find out that there is no significant difference along the treatments (Kruskall-Wallis test; $p = 0.55$). The same is true if we compare the average right transfer according to both the Recipients and the Observers along the treatments (Kruskall -Wallis test, $p = .15$; Mann-Whitney test, $p = 0.22$).

However, when we compare different participants' normative beliefs, it turns out that the Recipients' right transfer is significantly higher than the Dictators' and the Observers' ones (ttest, $p = 0.0002$ and $p = 0.0038$ respectively). On the other hand, Dictators' and Observers' beliefs are aligned (ttest, $p = 0.69$).

Result 4. *Subjective unfairness*

During the experiment the Observers were asked to identify their ideal transfer from the Dictator to the Receiver (see Table 3). If we assume that this ideal transfer is a subjective reference point of fairness, each Dictator's transfer lower than the ideal transfer

may be considered unfair on the part of the Observer. In this case, it is possible to analyse the Observer's reaction when her subjective principle of fairness is violated.

What turns out is that the Observers' levels of punishment are sensitive to their subjective sense of fairness. In Figure 2 the relation between the (subjectively perceived) unfairness of Dictators and the level of punishment from the Observer to the Dictator is depicted. It is clear that the level of punishment increases as the Dictator's transfer becomes lower and lower than the Observer's ideal transfer. However, some punishment still exists even when the Dictator transfers to the Receiver a sum which is higher than the Observer's ideal transfer⁶.

A random-effects Tobit regression of punishment on the variables *Negative Subjective Unfairness*⁷ and *Positive Subjective Unfairness*⁸ confirms the existence of a positive relation between the level of punishment and the degree of negative subjective unfairness ($p = 0.000$) as well as the negative relation between the level of punishment and the degree of positive subjective unfairness ($p = 0.08$). Moreover, it emerges that the Observer's reaction is significantly stronger in the MT when the Dictator's transfer becomes lower and lower than the Observer's ideal transfer ($p = 0.000$).

Subjective unfairness seems to be relevant also when we analyse participants A's behaviour. If we compare the Dictators' actual transfers and their ideal transfer, there is no significant difference (ttest, $p = 0.33$).

6 Discussion

Third Parties are willing to nonstrategically punish Dictators

Our analysis confirms one of Fehr and Fischbacher's (2004) major findings: most of third parties indeed punish 'unfair' Dictators and the amount of punishment is negatively related to the level of Dictators' transfers (Result 1). Hence, we confirm that the notion of strong negative reciprocity extends to the sanctioning behaviour of 'unaffected' third parties. Like

⁶ Such 'odd' punishment of 'hyper-fair' dictators qualitatively parallels the (often substantial amount of) so called 'perverse' punishment directed at high contributors discovered in virtually all the experiments based on the voluntary contribution mechanism (VCM; see e.g. Cinyabuguma et al., 2006).

⁷ *Negative Subjective Unfairness* is defined as $\max \{0, \text{Ideal Transfer} - \text{Actual Transfer}\}$

⁸ *Positive Subjective Unfairness* is defined as $\max \{0, \text{Actual Transfer} - \text{Ideal Transfer}\}$

Fehr and Fischbacher (2004), and unlike several experimental designs, we focus on one-shot interactions, so that punishment in the TPGT cannot be strategically motivated by the desire to induce higher contributions from other subjects in later periods. In other words, our design allows us to isolate nonstrategic sanctioning.

Observers' punishment is both nonstrategic and strategically motivated

With regard to the nature of Observers' punishment in our experiment, it is interesting to ask the following question: is their attitude towards punishment entirely nonstrategic? In principle it is not clear whether punishment behaviour underlies non-standard preferences or a relevant influence of emotional factors. While we cannot rule out that both factors are at work, our overall results tend to favour the first interpretation, as we find that Observers' punishment is partially strategically motivated. More specifically, we can reject the hypothesis that punishing subjects are driven by non-rational factors (such as anger) only, since our players acting as Observers in the MT are sensitive to the presence of potential punishers (i.e. the Receivers). In particular, we obtain this finding insofar as we take what we call 'subjective unfairness' into account (see Result 4). This result appears to be in line with the more general finding obtained so far by experimental studies on sanctioning, that is the ordinary good nature of punishment (see e.g. Anderson and Putterman, 2006).

Players share a common norm of fairness such as 'selfishness aversion', rather than an 'egalitarian' one

Our results show that, contrary to what others have hypothesized, when the experimental game takes the form of the well-known Dictator Game, experimental subjects do not see as 'fair' the so called 'egalitarian distribution norm' (see Result 4). While we confirm the previous finding that subjects are (altruistically) willing to enforce a norm of fairness even though the enforcement is costly for them (see on this Fehr and Fischbacher, 2004), a subjectively perceived norm of fairness (i) exists, (ii) is to some extent shared by all the players across treatments but (iii) it is different from a purely egalitarian norm of fairness. Fehr and Fischbacher (2004) hypothesize that the salient distribution norm in the Dictator

Game is for A to transfer half of the ‘pie’ to B, arguing that since the players interact anonymously and are randomly assigned their roles, there is no reason why A should end up with more money than B. By contrast, by eliciting players’ normative beliefs, we find that all of them expect a Dictator to give *something*, not to give half of his endowment to the Receiver. In our view, all this means that in all treatments Dictators, Observers and Receivers believe that, if you happen to play as a Dictator, playing entirely selfishly is unfair. However, we interestingly also see that fairness is subjectively perceived by all of them not egalitarianly but as *selfishness aversion*: while all the players agree that Dictators should avoid to keep the whole amount for themselves, they do not believe that splitting the pie equally is morally compulsory.

Appendix

Table 1.

Average Punishment in ...	When the Dictator transfers ...					
	0	1	2	3	4	5
TPGT	2	1.75	1.35	1.3	0.95	0.85
MT	2.05	1.79	1.37	0.89	0.68	0.58
<i>Mann-Whitney test</i>	$p = 0.56$	$p = 0.48$	$p = 0.32$	$p = 0.14$	$p = 0.26$	$p = 0.72$

Table 2.

	A transfers to B (1)	A thinks it is right to transfer to B (2)	B thinks it is right to transfer to B (3)	C thinks it is right to transfer to B (4)	<i>Kruskall - Wallis test</i> (2) = (3) = (4)	<i>t- test</i> (2) = (3) (3) = (4) (2) = (4)
DGT	1.4	1.5	3.1			$p = 0.0002$ $p = 0.0038$ $p = 0.69$
TPGT	1.55	1.85	3	2	$p = 0.11$	
MT	1.53	1.42	2.1	1.42	$p = 0.35$	
<i>Kruskall - Wallis test</i> <i>DGT = TPGT = MT</i>	$p = 0.88$	$p = 0.55$	$p = 0.15$	$p = 0.22$		

Table 3.

Percentage of cases of		0	1	2	3	4	5
DGT	Dictators who transfer...	40%	15%	25%	10%	5%	5%
	Dictators who think the right transfer is ...	30%	20%	30%	15%	0%	5%
	Recipients who think the right transfer is...	5%	10%	30%	15%	5%	35%
	Observers who think the right transfer is ...	-	-	-	-	-	-
TPGT	Dictators who transfer...	30%	20%	25%	20%	0%	5%
	Dictators who think the right transfer is ...	30%	10%	25%	25%	0%	10%
	Recipients who think the right transfer is ...	10%	0%	40%	15%	0%	35%
	Observers who think the right transfer is ...	25%	5%	30%	30%	5%	5%
MT	Dictators who transfer...	21%	42%	21%	0%	11%	5%
	Dictators who think the right transfer is ...	32%	26%	32%	0%	0%	11%
	Recipients who think the right transfer is ...	26%	11%	32%	11%	0%	21%
	Observers who think the right transfer is ...	47%	11%	16%	11%	11%	5%

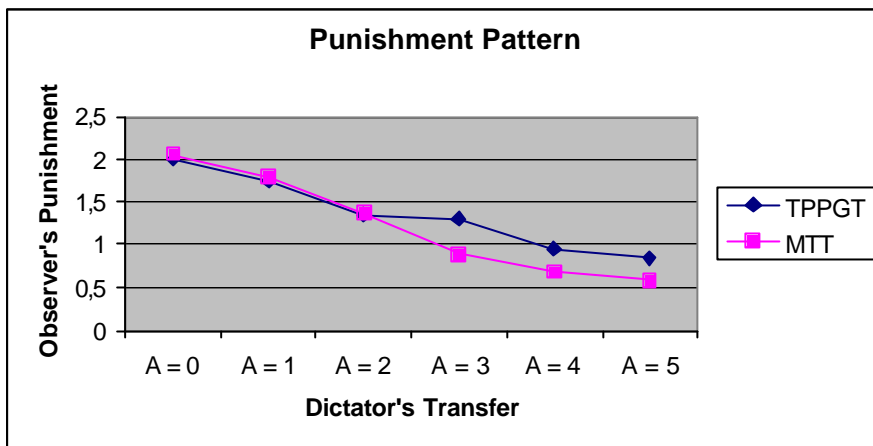


Fig. 1 Observer's behaviour

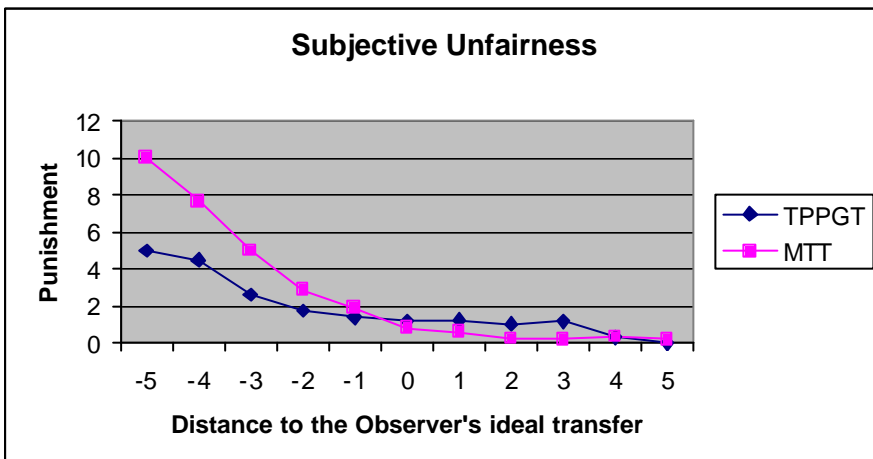


Fig. 2 Subjective Unfairness

References

- Anderson C.M., Putterman L., 2006, Do Non-Strategic Sanctions Obey the Law of Demand? The Demand for Punishment in the Voluntary Contribution Mechanism, *Games and Economic Behavior*, 54, pp. 1-24.
- Andreoni J., Miller J., 2002, Giving according to GARP: An experimental test of the consistency of preferences for altruism, *Econometrica*, 70, pp. 737-753.
- Axelrod R., 1986, An Evolutionary Approach to Norms, *The American Political Science Review*, 80, pp. 1095-1111.
- Bernhard H., 2005, 'Third Party Punishment within and across Groups – An Experimental Study in Papua New Guinea', Institute for Empirical Research in Economics, University of Zürich, Preliminary Working Paper.
- Bolton G.E., Ockenfels A., 2000, 'A Theory of Equity, Reciprocity and Competition', *American Economic Review*, 90, pp. 166-93.
- Carpenter J., 2007, The demand for punishment, *Journal of Economic Behavior and Organization*, 62, pp. 522-542.
- Cinyabuguma M., Page T., Putterman L., 2006, Can Second-order Punishment Deter Perverse Punishment, *Experimental Economics*, 9, pp. 265-279.
- Charness G., Rabin M., 2002, 'Social Preferences: Some Simple Tests and a New Model', *Quarterly Journal of Economics*, 117, pp. 817-69 .
- Eckel C.C., Grossman P.J., The Relative Price of Fairness: Gender Differences in a Punishment Game, *Journal of Economic Behavior and Organization*, 30, pp. 143-158.
- Egas M., Riedl A., 2005, The Economics of Altruistic Punishment and the Demise of Cooperation, Tinbergen Institute Discussion Paper.
- Fehr E., Fischbacher U., 2004, Third-Party Punishment and Social Norms, *Evolution and Human Behavior*, 25, pp. 63-87.
- Fehr E., Gächter S., 2000, Cooperation and Punishment, *American Economic Review*, 90, pp. 980-994.
- Fehr E., Schmidt K., 1999, 'A Theory of Fairness, Competition and Cooperation', *Quarterly Journal of Economics*, CXIV, pp. 817-51.
- Fischbacher U., 2007, zTree: Zurich Toolbox for Ready-made Economic Experiments, *Experimental Economics*, 10, pp. 171-178.

Gintis H., Bowles S., Boyd R., Fehr E., 2003, Explaining Altruistic Behavior in Humans, *Evolution and Human Behavior*, 24, pp. 153-172.

Guth W., Schmittberger R., Schwartz B., 1982, An Experimental Analysis of Ultimatum Games, *Journal of Economic Behavior and Organization*, 3, pp. 367-388.

Ottone S., 2005, Transfers and Altruistic Punishments in Solomon's Game Experiments, Paper n.57, AL.EX Series, POLIS, University of Eastern Piedmont.

Ottone S., 2007, Are People Samaritans or Avengers?, Paper n.86, AL.EX Series, POLIS, University of Eastern Piedmont.

Sally D., 1995, Conversation and Cooperation in Social Dilemmas: A Meta-Analysis of Experiments, *Rationality and Society*, 7, pp. 58-92.

Suleiman R., 1996, Expectations and Fairness in a modified Ultimatum Game, *Journal of Economic Psychology*, 17, pp. 531-554.

Zwick R., Chen Z.P., 1999, What Price for Fairness? A Bargaining Study, *Management Science*, 45, pp. 804-823.

Survey Evidence on Conditional Norm Enforcement

Christian Traxler* and Joachim Winter†

This version: January 23, 2008

Abstract

We discuss survey evidence on individuals' willingness to sanction norm violations such as fare dodging, speeding, or drunk driving. The data suggest that individuals are more prone to sanction norm violations which they believe to be rare. This pattern is in line with models of social norms and conditional cooperation.

JEL classification: K42; Z13; D1

Keywords: Norm Enforcement; Sanctioning; Survey Evidence; Social Norms.

* *Corresponding Author.* Max Planck Institute for Research on Collective Goods. Kurt-Schumacher-Str. 10, 53113 Bonn, Germany. Phone/Fax: +49 (0)228 91416-69/-62; E-mail: traxler@coll.mpg.de

† Department of Economics, University of Munich, Ludwigstr. 28 RG, 80539 Munich, Germany. Phone/Fax: +49 (0)89 2180-2459/-3954; E-mail: winter@lmu.de

1 Introduction

Over the last decade, a solid body of empirical research has documented the prevalence of conditional cooperation (see e.g. Fischbacher et al., 2001; Gächter, 2006, provides a comprehensive survey). Individuals who follow this behavioral pattern condition their cooperation – or their pro-social conduct in general – on (their belief about) the other’s inclination to act pro-socially. For instance, conditional cooperators would donate the more to charity, the more they think others give (Frey and Meier, 2004).

Conditionally pro-social behavior can be explained by the presence of social norms, i.e. rules of conduct which are enforced by internal or external sanctions (Coleman, 1990).¹ An important assumption of the literature is that these sanctions depend on the level of norm adherence: The more people comply with a norm, the stronger are the sanctions for a norm deviation (see e.g. Lindbeck et al., 1999). Given this pattern, social norms can induce conditional norm compliance.

The present paper discusses survey evidence which confirms this sanctioning pattern. In a national survey conducted in Austria, respondents were confronted with eight different “incorrect behaviors”, including drunk driving, speeding, fare dodging or skiving off work. Respondents were then asked how they would react if an acquaintance followed such behavior. Among the response options, two – cooling down the contact and expressing disapproval – represent norm enforcing sanctions which are well studied in experimental economics (see e.g. Masclet et al., 2003; Cinyabuguma et al., 2005). We show that sanctions typically follow a conditional pattern: People are more prone to sanction a norm violation, the more rarely it is believed to occur. This pattern is particularly pronounced among individuals with higher levels of education. We further find that those norm deviations which are sanctioned most frequently are sanctioned independently of beliefs about its pervasiveness. Hence, sanctions for the violation of such ‘strong norms’ seem to be unconditional.

2 The Survey

Our data stem from a study on TV licence fee compliance conducted by the Austrian Public Broadcasting Company in 2000. The survey was administered by GfK, a commercial survey

¹See Fehr and Schmidt (2006) for a survey of different models that can explain conditional cooperation.

organization, using computer-assisted personal interviewing. It was targeted at a random sample of the Austrian household population. From each household, the financial decision maker (defined as the one who is in charge of paying regular household expenditures like the rent or the electricity bill) was asked to answer the survey. The top panel of table 1 contains the socioeconomic characteristics of the respondents. To facilitate interpretation, all characteristics are grouped and coded as dummies. The sample is composed by 59% of females; slightly more than half of the respondents are older than 50 years. 55% have a college education, 19% hold a high-school or university diploma. 44% state that they have infrequent contact with their neighbors, every fifth has frequent contact.

Table 1 about here.

In one section of the survey, respondents were confronted with eight different behaviors ('acts') which were described as "*all incorrect but nevertheless occurring more or less frequently.*"²

- (1) *Drunk Driving*. Driving a car although one is aware of the fact that one has drunken too much and is clearly above the legal blood alcohol limit.
- (2) *Hazardous Waste*. Putting hazardous waste like batteries or chemicals into the ordinary household rubbish (instead of a waste separation container).
- (3) *Speeding*. Driving at 160kmph (instead of 120kmph) on a highway or at 70kmph (instead of 50kmph) in a residential area.
- (4) *Stealing Newspapers*. Stealing newspapers from pay-as-you-take news boxes.
- (5) *Skiving off Work*. Pretending to be sick and stay home from work for one or two days.
- (6) *Cheating on TV Licence Fees*. Not registering broadcasting equipment for TV licence fees.
- (7) *Fare Dodging*. Travelling on public transport without a ticket.
- (8) *Shadow Economy*. Hiring craftsmen from the shadow economy (without paying taxes).

²The original sequence of the behaviors in the survey was 2, 4, 7, 6, 8, 5, 3, 1.

All these behaviors constitute violations of Austrian laws.³ Respondents were asked to state their belief about how widespread these behaviors are on a five-point Likert scale ranging from ‘very infrequent’ [1] to ‘very frequent’ [5]. The distributions of the responses are presented in the lower panel of table 1. The data shows that the beliefs differ both, between individuals and between the different norm violations. For example, drunk driving and speeding (acts 1 and 3) are perceived to be more widespread than the evasion of TV licence fees (act 6).

The interviewed persons were then asked how they would react if they learned that a close acquaintance has taken act 1, 2 and so on. The five response categories,

[1] “*I would be impressed by him/her*”

[2] “*She/he should better not be caught*”

[3] “*I would not care*”

[4] “*I would seriously talk with him/her about this behavior and would try to convince him/her to stop doing it*”

[5] “*I would cool down the contact with him/her*”

can be ordered from approval [1], (benevolently) ignoring it [2, 3] to sanctions in the form of expressing disapproval [4] or even exclusion [5]. The distributions of the responses are reported in table 2.

Table 2 about here.

A majority would sanction an acquaintance for the first five acts – mostly by expressing disapproval rather than by cooling down the contact. More than 80% would sanction the first two acts, still 66% would do so for act 3. Hence, the social norms against these three norm violations – which are also those with the (potentially) most grave externalities – are particularly *strong*, in the sense that at least two-third of the respondents are willing to sanction it. Equivalently,

³Austrian citizens are bound by law to bring hazardous waste to waste separation containers (act 2). On Sundays, Austrian newspapers can be bought from pay-as-you-take boxes; these boxes, however, have no lock which would prevent people from stealing the paper (act 4). (For a related field experiment, see Pruckner and Sausgruber, 2006.) In case of sickness, Austrian employees are allowed to stay at home for two days without any medical certificate of their sickness (act 5). Owners of any broadcasting equipment have to register their equipment and pay an annual licence fee of roughly 200 Euros at the time of the survey (act 6).

one might call the social norms against acts 6–8 *weak*, since it is only sanctioned by a minority: 50% and more would react with ignorance. Finally, note that none of the eight behaviors meets with a significant share of approval.

3 Conditional Norm Enforcement

How does sanctioning differ between individuals and how does it relate to their beliefs about the frequency of the different acts? We answer these questions by estimating ordered probit models. We assume that respondent i has propensity \tilde{y}_{ij} to sanction act $j \in \{1, \dots, 8\}$. These propensities are latent variables which are unobserved in our data. The observed dependent variables, the individual reactions $y_{ij} \in \{1, \dots, 5\}$, are determined by the response model

$$y_{ij} = \begin{cases} 1 & \text{if } \tilde{y}_{ij} \leq \gamma_{1j} \\ 2 & \text{if } \gamma_{1j} < \tilde{y}_{ij} \leq \gamma_{2j} \\ \dots & \\ 5 & \text{if } \gamma_{4j} < \tilde{y}_{ij} \end{cases}$$

and $\tilde{y}_{ij} = \alpha_j Bel_{ij} + \beta_j X_i + \varepsilon_{ij}$, where the vector X_i contains the socioeconomic characteristics (see table 1) and ε_{ij} is a random component which is assumed to be i.i.d. normal across respondents i and acts j , conditional on the covariates.

Our key explanatory variable is Bel_{ij} , the individuals' belief about the frequency with which the different acts occur. Remember that higher values of \tilde{y}_{ij} represent a stronger propensity to sanction and that higher values of Bel_{ij} mean that act j is expected to be more common. Hence, if the severity of sanctions are declining (and accordingly, reactions are more positive) the more often a norm is violated – as it is assumed in the literature on social norms (see e.g. Lindbeck et al., 1999) – we should expect negative coefficients for α_j .

The results from our estimations are given in table 3. The left part of the table presents the outcome for the full sample.⁴ Except for one behavior (act 2, hazardous waste), α_j has the expected negative sign. The coefficients, however, are only significant for the last five

⁴Item non-response differs somewhat between the eight acts (see also table 2). We estimated alternative model specifications which control for non-response on the belief question. These estimations showed that our findings are not affected by the non-responses. We also estimated binary probit models explaining whether individuals sanction (choose reaction 4 or 5). The results were equivalent to those presented here.

norm violations (acts 4–8). For deviations from the three strong norms – the most frequently sanctioned acts 1–3 – we do not find a significant impact of the individuals’ belief.

Table 3 about here.

In a next step, we replicate the estimations for a restricted sample which excludes individuals with low education. The results are reported in the right part of table 3. For deviations from the two strongest norms, we find a weakly significant conditional sanctioning behavior for act 1 (drunk driving) and no significant effect for act 2 (hazardous waste). For the remaining six acts there is a clear conditional sanctioning pattern: The less frequent these norm deviations are expected to occur, the more severely they are sanctioned. This effect is significant at the 1% (acts 3, 4, 6, 7 and 8) and 5%-level (act 5), respectively.

The response to deviations from strong norms (acts 1–3) seems to be unconditionally negative, whereas sanctions for violations of weaker norms are conditioned on the expected prevalence of the respective acts. Table 3 also suggests that this conditional sanctioning pattern is more pronounced among the population with intermediate and high levels of education. This finding is consistent with the observation that conditional cooperation is particularly robust in lab experiments with student subject pools (see Gächter, 2007). The data further show that females tend to be more inclined to sanction, in particular deviations from the strong norms. In contrast, employed respondents are less engaged in sanctioning. All other socioeconomic characteristics do not have a clear-cut impact on sanctioning behavior.

4 Concluding Discussion

The evidence from a national survey presented in this paper supports a central assumption of the literature on social norms: The strength of norm enforcing sanctions depends on the expected level of norm compliance. The more people deviate from a social norm, the less severely a norm violation is sanctioned. This pattern is most pronounced for ‘weaker’ norms and among individuals with higher education.

All eight behaviors which are considered in the survey correspond to law violations. Hence, our findings have also implications for the literature on law and economics. Research within

this field has explained high compliance with ‘mild’ laws – laws which are only backed by minor legal sanctions – by the fact that people are also governed by social norms: legal *and* informal sanctions together contribute to compliance (Posner, 2000). If the informal sanctions for the violation of a (legal and social) norm are conditional on the (expected) prevalence of such acts – as suggested by our evidence – it establishes a rationale for condition compliance with laws. In this way, our results are compatible with multiple equilibrium states in criminal activity – in particular such petty crimes as stealing newspapers or fare dodging (compare e.g. Glaeser et al., 1996).

Acknowledgements

Financial support by the Austrian National Bank (*OeNB Jubiläumsfonds* Grant No. 12301) as well as the Marie Curie Research Training Network ENABLE is acknowledged. Simon Gächter and Roberto Galbiati provided helpful comments.

References

- Coleman, James S. (1990), *Foundations of Social Theory*, Harvard University Press, Cambridge (MA).
- Cinyabuguma, Matthias, Talbot Page and Louis Putterman (2005), Cooperation under the threat of expulsion in a public goods experiment, *Journal of Public Economics* 89(8), 1421-1435.
- Fehr, Ernst and Klaus Schmidt (2006), The Economics of Fairness, Reciprocity and Altruism – Experimental Evidence and New Theories, in: Serge-Christophe Kolm and Jean Mercier Ythier (Eds.), *Handbook on the Economics of Giving, Reciprocity and Altruism*, Vol.1, North Holland, Amsterdam.
- Fischbacher, Urs, Simon Gächter and Ernst Fehr (2001), Are People Conditionally Cooperative? Evidence from a Public Goods Experiment, *Economics Letters* 71(3), 397-404.
- Frey, Bruno and Stephan Meier (2004), Social Comparisons and Pro-social Behavior: Testing

‘Conditional Cooperation’ in a Field Experiment, *American Economic Review* 94(5), 1717-1722.

Gächter, Simon (2007), Conditional cooperation: Behavioral regularities from the lab and the field and their policy implications, in: Bruno S. Frey and Alois Stutzer (Eds.): *Economics and Psychology*, MIT Press, Cambridge (MA).

Glaeser, Edward L., Bruce Sacerdote and Jose A. Scheinkman (1996), Crime and Social Interaction, *The Quarterly Journal of Economics* 106, 507-548.

Lindbeck, Assar, Sten Nyberg and Jörgen W. Weibull (1999), Social Norms and Economic Incentives in the Welfare State, *The Quarterly Journal of Economics* 114(1), 1-35.

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), 366-380.

Posner, Richard A. (2000), *Law and Social Norms*, Harvard University Press, Cambridge (MA).

Pruckner, Gerald and Rupert Sausgruber (2006), Trust on the Streets: A Natural Field Experiment on Newspaper Purchasing, Discussion Papers, University of Copenhagen.

Tables

Table 1: Descriptive Statistics for the Covariates

<i>Variable</i>	N	Mean	SD	Min	Median	Max
Female	1138	0.59	0.49	0.00	1.00	1.00
Employed	1135	0.50	0.50	0.00	0.00	1.00
Age						
Age Low (< 29)	1138	0.09	0.29	0.00	0.00	1.00
Age Mid (30–49)	1138	0.39	0.49	0.00	0.00	1.00
Age High (> 50)	1138	0.52	0.50	0.00	1.00	1.00
Education						
Edu Low	1138	0.27	0.44	0.00	0.00	1.00
Edu Mid	1138	0.55	0.50	0.00	1.00	1.00
Edu High	1138	0.19	0.39	0.00	0.00	1.00
Household Net Income						
Inc Low (< 1017€)	1114	0.13	0.34	0.00	0.00	1.00
Inc Mid (1017–2024€)	1114	0.47	0.50	0.00	0.00	1.00
Inc High (> 2024€)	1114	0.40	0.49	0.00	0.00	1.00
Contact with Neighbors						
Con Low	1136	0.44	0.50	0.00	0.00	1.00
Con Mid	1136	0.35	0.48	0.00	0.00	1.00
Con High	1136	0.20	0.40	0.00	0.00	1.00
Belief about Frequency						
Drunk Driving	1118	3.59	1.01	1.00	4.00	5.00
Hazardous Waste	1095	3.34	1.23	1.00	3.00	5.00
Speeding	1123	4.17	0.91	1.00	4.00	5.00
Stealing Newspapers	1030	3.16	1.25	1.00	3.00	5.00
Skiving of Work	1074	3.09	1.17	1.00	3.00	5.00
Licence Fees	1052	2.64	1.09	1.00	3.00	5.00
Fare Dodging	984	2.96	1.15	1.00	3.00	5.00
Shadow Economy	1094	3.81	1.05	1.00	4.00	5.00

Table 2: Reaction to an Acquaintance’s Norm Violation (in Percentages)

	Drunk Driving	Hazard. Waste	Speeding	Stealing Newsp.	Skiving of Work	Cheat on TV Fees	Fare Dodging	Shadow Economy
Would freeze contact	4.39	7.82	5.98	13.27	11.69	8.00	9.31	7.29
Seriously talk about it	78.65	73.11	59.40	42.88	40.07	38.22	34.36	13.27
Would not care	9.58	12.30	20.21	31.11	34.80	40.16	34.71	50.09
Better not be caught	5.54	4.04	12.30	9.75	9.40	9.58	15.55	21.70
Would be impressed	0.09	0.09	0.18	0.26	0.79	0.53	0.09	4.04
No Response	1.76	2.64	1.93	2.72	3.25	3.51	5.98	3.60

Table 3: Ordered Probit Estimations - Dependent Variable: Reaction

	<i>Full Sample</i>								<i>Restricted Sample (Intermediate & High Education)</i>							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Drunk Driving	Hazard.Waste	Speeding	Steal.Newsp.	Skiving of Work	TV Fees	Fare Dodging	Shadow Econ.	Drunk Driving	Hazard.Waste	Speeding	Steal.Newsp.	Skiving of Work	TV Fees	Fare Dodging	Shadow Econ.
Belief	-0.047	0.03	-0.059	-0.095	-0.081	-0.078	-0.062	-0.204	-0.087	-0.006	-0.119	-0.115	-0.086	-0.113	-0.109	-0.234
	[1.22]	[0.94]	[1.52]	[3.19]***	[2.75]***	[2.42]**	[1.88]*	[6.17]***	[1.91]*	[0.17]	[2.58]***	[3.33]***	[2.49]**	[3.12]***	[2.93]***	[5.93]***
Female	0.19	0.118	0.156	0.026	0.063	0.018	-0.056	0.127	0.255	0.188	0.223	0.027	0.03	0.07	-0.01	0.119
	[2.35]**	[1.50]	[2.19]**	[0.36]	[0.89]	[0.26]	[0.77]	[1.78]*	[2.63]***	[2.06]**	[2.67]***	[0.33]	[0.38]	[0.86]	[0.12]	[1.47]
Employed	-0.164	-0.219	-0.177	-0.223	0.09	-0.224	-0.26	-0.039	-0.189	-0.314	-0.215	-0.213	0.033	-0.231	-0.288	-0.048
	[1.65]*	[2.31]**	[1.93]*	[2.50]**	[1.09]	[2.74]***	[2.84]***	[0.47]	[1.58]	[2.90]***	[2.02]**	[2.12]**	[0.35]	[2.57]**	[2.81]***	[0.50]
Con High	-0.086	-0.043	-0.122	0.033	0.036	0.081	-0.013	0.098	-0.092	-0.088	-0.031	0.002	0.065	0.021	-0.046	0.132
	[0.79]	[0.41]	[1.29]	[0.34]	[0.38]	[0.83]	[0.13]	[1.06]	[0.74]	[0.72]	[0.28]	[0.02]	[0.62]	[0.19]	[0.40]	[1.25]
Con Low	0.079	0.055	0.152	0.087	0.046	0.034	0.064	0.064	0.132	0.064	0.248	0.085	0.012	0.024	0.131	0.116
	[0.88]	[0.66]	[1.92]*	[1.09]	[0.60]	[0.45]	[0.78]	[0.82]	[1.21]	[0.64]	[2.66]***	[0.91]	[0.13]	[0.27]	[1.37]	[1.25]
Age High	0.038	-0.05	0.173	0.113	0.237	0.093	0.055	0.089	0.024	-0.096	0.086	0.209	0.18	0.117	0.09	0.065
	[0.38]	[0.50]	[1.87]*	[1.23]	[2.75]***	[1.08]	[0.58]	[1.07]	[0.20]	[0.83]	[0.82]	[2.02]**	[1.83]*	[1.24]	[0.86]	[0.68]
Age Low	0.188	-0.062	-0.107	-0.094	-0.044	-0.116	-0.066	-0.045	0.271	0.07	-0.113	-0.037	-0.065	-0.081	-0.027	-0.139
	[1.63]	[0.44]	[0.78]	[0.79]	[0.34]	[0.88]	[0.50]	[0.35]	[1.91]*	[0.44]	[0.72]	[0.28]	[0.45]	[0.57]	[0.19]	[0.96]
Inc High	0.023	0.01	0.045	0.048	-0.013	-0.02	-0.005	0.046	0.005	-0.016	0.062	0.091	-0.011	-0.023	-0.062	-0.016
	[0.26]	[0.12]	[0.58]	[0.63]	[0.18]	[0.27]	[0.06]	[0.62]	[0.05]	[0.17]	[0.71]	[1.08]	[0.13]	[0.28]	[0.75]	[0.19]
Inc Low	-0.029	0.021	0.124	-0.095	-0.038	0.048	0.021	0.013	0.037	0.011	0.135	-0.026	0.118	0.198	-0.042	0.082
	[0.21]	[0.17]	[1.08]	[0.81]	[0.32]	[0.42]	[0.19]	[0.11]	[0.18]	[0.06]	[0.80]	[0.14]	[0.67]	[1.13]	[0.27]	[0.49]
Edu High	0.15	-0.072	0.062	-0.095	0.135	-0.124	-0.152	0.209	0.141	-0.067	0.047	-0.11	0.124	-0.122	-0.114	0.236
	[1.60]	[0.74]	[0.67]	[0.99]	[1.49]	[1.39]	[1.58]	[2.41]**	[1.33]	[0.65]	[0.47]	[1.13]	[1.32]	[1.33]	[1.19]	[2.64]***
Edu Low	-0.014	-0.016	0.045	0.159	-0.062	0.057	0.099	0.151								
	[0.13]	[0.16]	[0.46]	[1.72]*	[0.69]	[0.62]	[1.05]	[1.71]*								
N	1077	1052	1080	988	1028	1010	937	1041	798	784	799	744	771	750	717	773

Robust z Statistics in Squared Brackets; * / ** / *** indicate Significance at a 10% / 5% / 1%-Level

All estimations include region dummies (estimated coefficients not reported).

When Equality Trumps Reciprocity: Evidence from a Laboratory Experiment

Erte Xiao^{†‡} and Cristina Bicchieri*

Research in economics, psychology and sociology provides compelling evidence that people often make decisions inconsistent with monetary earnings-maximization. From this observation has emerged a substantial empirical and theoretical literature seeking to improve our understanding of human decision making. This literature points to inequity aversion and reciprocity as key important motivations underlying actual human decisions (see, Fehr and Gächter, 2000). However, these two motivations do not always offer convergent implications for decisions. In particular, if a beneficiary is less wealthy than his benefactor then his reciprocal action would necessarily increase inequality. Given that inequality is a ubiquitous feature of human society in general, and that it typically exists in principle-agent relationships in particular, it is perhaps surprising that very little previous research has attempted to characterize decision making when equality and reciprocity are irreconcilable. This paper provides, to our knowledge, the first direct evidence on decision making in environments where inequality aversion and reciprocity have divergent behavioral implications.

Humans are averse to inequitable outcomes (Walster, Walster and Berscheid, 1978), and are willing to incur costs to reduce inequality between themselves and their counterparts (see. Güth et al. 1982). Importantly, this is true regardless of the source of the inequality: behavior motivated by equity is independent of intentions (e.g. Dawes, et al, 2007).

On the other hand, substantial evidence reveals that inequality aversion cannot fully explain behavior, and intentions do matter (e.g. Blount, 1995; Charness and Haruvy, 2002; Falk, Fehr and Fischbacher, forthcoming; Cox, 2004). For example, people might respond more negatively when an unequal outcome is caused intentionally by a person than when it is brought about by nature. Positive reciprocity refers to a positive reaction to kind intentions. Negative reciprocity refers to negative actions in response to hostile intentions, even when these acts yield no future or current monetary payoffs and might even be costly (see Fehr and Gächter, 2000 for an excellent discussion of reciprocity).

Reciprocity relationships between principals and agents have been widely studied. Agents reciprocating to the good intentions of principals have been shown to play an important role in promoting cooperative relationships in economic exchange (see, Berg, Dickhaut and McCabe, 1995). On the other hand, a principal (e.g. an employer) might very often be wealthier than an agent, so that reciprocation would also increase the wealth or earnings inequality between them. This could potentially reduce one's propensity to reciprocate and consequently undermine trust. Thus, to understand how people behave when the equality motives conflict with preferences for reciprocity is evidently important when attempting to design institutions to promote cooperation.

This paper focuses on the standard case where reciprocating to kind intentions (i.e. positive reciprocity) cannot be reconciled with inequality aversion. Our baseline experiment is similar to a standard trust game. An investor and a trustee are both given the same endowment. The investor can decide whether to transfer a certain amount to the trustee. The trustee receives the tripled transfer amount and decides whether to transfer back an amount to the investor. In this case, any reciprocal return from the trustee to the

investor that is less or equal than two-third of the tripled transfer amount also reduces inequality.

In our second treatment, an asymmetry treatment, the trustee receives the same endowment as in the baseline treatment, but the investor is given relatively more, and in such a way that both investor and trustee earn equal amounts in the event that the investor decides to transfer and the trustee returns zero. Thus, any positive return by the trustee increases inequality.

We find, in relation to a standard trust-game treatment where trustees' responses reduce inequality, the proportion of non-reciprocal decisions is twice as large when reciprocity promotes inequality. Moreover, investors expect that this will be the case. Overall, although both motives clearly play a role, more of our data can be explained by inequality aversion than by reciprocity. Our results call attention to the potential importance of inequality in principal-agent relationships, and have important implications for designing policies aimed at promoting cooperation.

[†] University of Pennsylvania, Philosophy, Politics and Economics Program and Wharton School, 313 Logan Hall, 249, S. 36th Street, Philadelphia, PA 19104, 215.746.3618 (office), exiao@sas.upenn.edu

^{*} University of Pennsylvania, Philosophy, Politics and Economics Program, Department of Philosophy and Wharton School, 466 Logan Hall, 249, S. 36th Street, Philadelphia, PA 19104, 215.898.5820 (office), cb36@sas.upenn.edu

[‡] Correspondence to E.X.

XI Summer School on Economics and Philosophy: SOCIAL NORMS

San Sebastian (Spain), 14-17 July 2008

INTENTIONALITY AND DECISION-MAKING IN CHILDREN: A RESEARCH WITH THE ULTIMATUM GAME

Antonella Marchetti*, Iliaria Castelli*, Alan G. Sanfey^o

**Research Unit on Theory of Mind, Department of Psychology,
Università Cattolica del Sacro Cuore, Milano, Italy*

*^o Neural Decision Science Laboratory,
University of Arizona, Tucson, USA*

Corresponding author: ilaria.castelli@unicatt.it

THEORETICAL BACK-GROUND

Decision-making constitutes a cross-current object of research among Economics, Psychology and Neuroscience (Sanfey, 2004; Sanfey et al., 2006) and has become considered a complex process involving both deliberative and emotional-affective components (Bechara et al., 1997; Camerer, 2003; Pillutla, Murnighan, 1996; Sanfey, 2007; Sanfey et al., 2003; van't Wout et al., 2006). The strongest evidences have been provided by intense research in the field of behavioural economics with various types of social interactive games involving monetary exchanges (see Camerer, 2003 for an overview). One of most used is the Ultimatum Game – UG – (Güth, Schmittberger, Schwarze, 1982): a proposer (P) makes an offer about the division of a certain amount of money and a responder (R) decides to accept (both P and R earn something) or to refuse (both P and R earn nothing). The results of numerous studies in western industrial countries have showed that people's behaviour in the UG systematically contradict the predictions of classical economic theories of ideally rational maximizing decision-makers; in fact, good and even equal splits are made by P and offers around 20-30% of the amount are rejected by R half of the times (Camerer, 2003). It is therefore clear that in social interactions involving money people do not care only for maximizing their payoffs, but pay attention also to other elements. For example, the social role they cover in such game: even a simple connotation of the players as “buyers” vs “sellers” (Hoffman et al., 1994) has shed light on how social roles and the expectations connected to those roles can affect the behaviour of players. Another key element is the sensibility to fairness (Fehr, Schmidt, 1999) or, more specifically, to a social norm for fairness (Bicchieri, 2006; Bicchieri, Chavez, 2007) which may be obeyed or not depending on the types of expectations that people have about other people following that norm. In other words, the behaviour in the UG can be explained referring to what players expect others to do in terms of fairness (empirical expectations) and to what they believe others think ought to be done in term of fairness (normative expectations).

Other psychological processes, such as emotions and mentalizing or theory of mind – the representation of one's own and other person's intentions and mental states (Premack, Woodruff, 1978) – are involved as well. Emotions play a mayor role especially on R's side, since her anger and spite for being treated unfairly make the refusal of unfair offers easily understandable in terms of “punishment” of the greedy P (Pillutla, Murnighan, 1996; Sanfey et al., 2003). Mentalizing contributes to understand the intra and inter-personal mentalistic representations of the intentions both on the side of P (Hoffman, et al., 2000) and on the side of R (Marchetti, Castelli, Sanfey, 2007). Moreover, the level of intentionality that R attributes to P – human vs non-human – plays a mayor role as well (Blount, 1995), inasmuch as the tendency to refuse unfair offers is connected to

offers that are perceived as intentionally unfair: unfair offers are rejected more if they are made by a human partner than by a computer, as showed by Rilling et al. (2004).

It is important to note that the majority of works about the UG – included the works quotes so far – have been carried on adults, whereas it would be interesting to understand how adults build those patterns of behaviour, i.e. to adopt a developmental perspective on decision-making. Some researches on decision-making in children and adolescents have been carried on indeed (Murnighan, Saxon, 1998; Harbaugh, Krause, Liday, 2003; Hoffman, Tee, 2006; Sally, Hill, 2006; Sutter, 2007) showing that generally young children tend to accept unfair offers more than old children and adolescents and that young children are inclined to make more generous offers than old children. In other words, the decision-making behaviour – both as P and as R – tends to become similar to those of adults as age grows.

RESEARCH

In this work we investigate the development of the decisional behaviour in children along with the acquisition of two other abilities:

(a) the understanding of intentions, i.e. the level of intentionality R attributes to P.

Sutter (2007) showed that intentionality matters less for children than for adults, since the younger children are, the more sensible they are to the outcome than to the partner's intentions;

(b) theory of mind, i.e. the ability to explain one's own and other people's behaviour as intentional by referring to mental states, an ability that undergoes major developmental changes during childhood (Wellman, Cross, Watson, 2002).

The questions that guide this research can be listed as follows:

- Is children's understanding of different levels of P intentionality associated to children's decisions in the UG? If so, in which of the UG offers (fair vs unfair)?
- Is children's understanding of theory of mind associated to their decisions in the UG about offers coming from intentional vs non-intentional proposers?

Participants are pupils of kindergarden schools and of Primary Schools in order to cover the age range from 4 years to 10 years. Children will be admitted at the research after parent's informed written consent and will be tested individually in a quiet room at school.

To ensure that the group of participants is homogenous in terms of cognitive abilities, Raven's Coloured Progressive Matrices – CPM (Raven, 1947) will be used and children with low scores (25th percentile or less) will be excluded from the analyses.

Children will be submitted to the UG (Marchetti, Castelli, Sanfey, 2008): they will play as responders for real with 10 tokens that will be converted into candies and/or stickers according to the child's preference. Each child will play many UG rounds some UG rounds with a human partner (attribution of intentionality) and some other UG rounds with a roulette (no attribution of intentionality), facing fair offers (5-5) and increasingly unfair offers (6-4, 7-3, 8-2, 9-1). The order (roulette vs human partner) will be counterbalanced across participants; the order of the fair/unfair offers will be set using the Latin square.

Finally, in order to evaluate theory of mind development, a first and second order false belief task (Marchetti, Sanfey, Castelli, 2006, revised version of the task by Antonietti, Liverta Sempio, Marchetti, 1999, compromise between Perner, Wimmer, 1985 and Sullivan et al., 1994) will be submitted. The child is told a story (with drawings) about Mary and John who are playing with a toy: Mary puts the toy in the wardrobe and leaves the room, while she is away John changes the location of the toy and puts it under the bed. The child is asked a 1st order false belief question ("Where will Mary look first for the toy when she comes back to the room?") and some control questions. Then the story starts again, with Mary coming back to the room: from the open door she sees John in the very moment he is changing the location of the toy, but John does not see Mary. The child is asked a 2nd order false belief question ("According to John, where will Mary look first for the toy when she comes back to the room?") and some control questions.

PRELIMINARY RESULTS

Data are being collected: preliminary results on children attending primary school show that in general young children (8 yrs.) tend to accept unfair offers more than old children (10 yrs.), in line with the literature on decision-making in children quoted before. Moreover, unfair offers are rejected more when they come from human partner: both 8 and 10 year olds behaviour is significantly associated to the type of P when offers become more unfair (8-2 and 9-1 offers are rejected more when they come from a human P), showing that children change behaviour according to the intentionality they attribute to P and that they are more sensible to P intentions than to the outcome when the offer is highly unfair.

Data acquisition from younger children (4-6 yrs.) should provide a more complete understanding of the developmental trend of decision-making behaviour along with the understanding of intentionality and of theory of mind. In fact results on 4 and 6 year old children shed light on an age-window that is critical for the relationship between decision-making and theory of mind, since the first order false belief reasoning is being acquired and well established and the second order false belief reasoning is under construction.

REFERENCES

Antonietti, A. Liverta Sempio, O., Marchetti, A. (1999). *I compiti di falsa credenza di secondo ordine di look-prediction e say-prediction*. Unità di Ricerca sulla Teoria della Mente, Dipartimento di Psicologia, Università Cattolica del Sacro Cuore, Milano.

Bechara, A., Damasio, H., Tranel, D., Damasio, A.R. (1997). Deciding advantageously before knowing the advantageous strategy. *Science*, 275,1293–1295.

Bicchieri, C. (2006). *The grammar of society. The nature and dynamics of social norms*. New York: Cambridge University Press.

Bicchieri, C., Chavez, A. (2007). *Behaving as Expected: Public Information and Fairness Norms*. Goldstone Research Unit Working Paper, December 2007.

Blount, S. (1995). When social outcomes are not fair: the effects of causal attributions on preferences. *Organizational Behavior and Human Decision Processes*, 63(2), 131-144.

Camerer, C.F. (2003). *Behavioral game theory*. New York: Russell Sage Foundation.

Fehr, E., Schmidt, K.M. (1999). A Theory of Fairness, Competition, and Cooperation. *The Quarterly Journal of Economics*, 114(3), 817-868.

Güth, W., Schmittberger, R., Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior and Organization*, 3, 367-388.

Harbaugh, W.T., Krause, K., Liday, S.G. (2003). Bargaining by children. <http://econpapers.repec.org/paper/oreuocwp/2002-04.htm>

Hoffman, E., McCabe, K., Shachat, S., Smith, V. (1994). Preferences, property rights and anonymity in bargaining games. *Games and economic behaviour*, 7, 346-380.

Hoffman, E., McCabe, K., Smith, V. (2000). The impact of exchange context on the activation of equity in ultimatum games. *Experimental Economics*, 3, 5-9.

Hoffman, R., Tee, J.Y. (2006). Adolescent-adult interactions and culture in the ultimatum game. *Journal of Economic Psychology*, 27, 98-116.

Marchetti, A., Castelli, I., Sanfey, A.G. (2007). *When and how ToM is involved in economic decision-making?* Oral communication presented at the 13th European Conference on Developmental Psychology (ECDP), Jena, Germany, August 21st-25th 2007 (CD-ROM abstracts).

Marchetti, A., Castelli, I., Sanfey, A.G. (2008). *Ultimatum Game for children with different proposers: roulette vs human partner*. Unità di Ricerca sulla Teoria della Mente, Dipartimento di Psicologia, Università Cattolica del Sacro Cuore, Milano e Neural Decision Science Laboratory, University of Arizona, Tucson.

Marchetti, A., Sanfey, A.G., Castelli, I. (2006). *“Look-prediction”- second Italian version*. Research Unit on Theory of Mind, Department of Psychology, Catholic University of the Sacred Heart, Milano (Italy) and Neural Decision Science Laboratory, University of Arizona, Tucson (USA).

- Murnighan, J.K., Saxon, M.S. (1998). Ultimatum bargaining by children and adults. *Journal of Economic Psychology*, 19, 415-445.
- Perner, J., Wimmer, H. (1985). 'John thinks that Mary thinks that...': attribution of second-order false beliefs by 5 to 10-year-old children. *Journal of Experimental Child Psychology*, 39(9), 437-471.
- Pillutla, M.M., Murnighan, J.K. (1996). Unfairness, anger, and spite: emotional rejections of ultimatum offers. *Organizational Behavior and Human Decision Processes*, 68(3), 208-224.
- Premack, D., Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *The Behavioural and Brain Science*, 1, 515 – 526.
- Raven, J.C. (1947). *Progressive Matrices : séries A, Ab, B*. Clamart Seine: Editions scientifiques et psychotechniques, 1951-1953.
- Rilling, J.K., Sanfey, A.G., Aronson, J.A., Nystrom, L.E., Cohen, J.D. (2004). The neural correlates of theory of mind within interpersonal interactions. *NeuroImage*, 22, 1694-1703.
- Sally, D., Hill, E. (2006). The development of interpersonal strategy: autism, theory of mind, cooperation and fairness. *Journal of Economic Psychology*, 27,73-97.
- Sanfey, A. G. (2004). Neural computations of decision utility. *Trends in Cognitive Sciences*, 8(12), 519-521.
- Sanfey, A. G. (2007). Social decision-making: insights from game theory and neuroscience. *Science*, 318, 598-602.
- Sanfey, A.G., Loewenstein, G., McClure, S.M., Cohen.J.D. (2006). Neuroeconomics: cross-currents in research on decision-making. *Trends in Cognitive Sciences*, 10(3), 108-116.
- Sanfey, A.G., Rilling, J.K., Aronson, J.A., Nystrom, L.E., Cohen, J.D. (2003). The neural basis of economic decision-making in the Ultimatum Game. *Science*, 300, 1755–1758.
- Sullivan, K., Zaitchik, D., Tager-Flusberg, H. (1994). Preschoolers can attribute second-order beliefs. *Developmental Psychology*, 30, 395-402.
- Sutter, M. (2007). Outcomes versus intentions: on the nature of fair behavior and its development with age. *Journal of Economic Psychology*, 28, 69–78.
- van 't Wout, M., Kahn, R.S., Sanfey, A.G., Aleman, A. (2006). Affective state and decision-making in the Ultimatum Game. *Experimental Brain Research*, 169, 564-568.
- Wellman, H.M., Cross, D., Watson, J. (2001). Meta-analysis of theory of mind development: the truth about false belief. *Child Development*, 72, 3, 655-684.

SOCIAL NORMS AND BELIEFS: AN EXPERIMENTAL INVESTIGATION

Marco Faillo^{*}
Stefania Ottone[^]
Lorenzo Sacconi[†]

Introduction. In the field of experimental economics one of the most studied topic is subjects' reaction when a cooperation norm or a redistribution norm is violated. This implies that the experimental literature concerning norms mainly corresponds to studies on fairness and, consequently, on punishment of defectors (f.i., Fehr and Gächter, 2000, for second-party punishment; Fehr and Fischbacher, 2004, for third-party punishment). A further implication of these mainstream experimental designs is that updating the classical figure of the *Homo Oeconomicus* by introducing social preferences (inequity aversion, reciprocity, altruism, spitefulness) into the economic theories is sufficient to explain the experimental results.

At the same time, models like Grimalda and Sacconi (2005), Sacconi and Grimalda (2007) and Bicchieri (2006) deal with the problem of the compliance with social norms, whose role is crucial when formal sanctions cannot be used and the reputational mechanism cannot be fully implemented as enforcement device. In both models the willingness to conform to shared social norms - which imply non-self-interested behaviour - depends on expectations about what people do (First Order Empirical Expectations – FOEE, and Second Order Empirical Expectations - SOEE) and/or what people expect ought to be done (Second Order Normative Expectations - SONE). In particular, according to Bicchieri's theory of social norms, in presence of a conflict of interests among agents conformity is conditional on both NE and EE.¹ In Sacconi and

* University of Trento

[^] EconomEtica and University of Eastern Piedmont

[†] University of Trento and EconomEtica

Grimalda's model compliance is the consequence of both of agents' participation in choosing the norm in a social contract setting and of the existence of expectations about others' willingness to conform (FOEE and SOEE).

The main contribution of this paper is twofold. First of all, it focuses on the decisional process that leads to the creation of a social norm. Secondly, it analyses the mechanisms through which subjects conform their behaviour to the norm. In particular, our aim is to study the role and the nature of Normative and Empirical Expectations and their influence on people's decisions.

Experimental Design. The tool is the *Exclusion Game* (Sacconi and Faillo, 2005; Faillo and Sacconi, 2007), a sort of 'triple mini-dictator game'. It represents a situation where 3 subjects – players A - have to decide how to allocate a sum S among themselves and a fourth subject - player B - who has no decisional power. In particular, each player A has to decide the amount s/he wants to ask for her/himself choosing one of three possible strategies: asking 25%, 30% or 33% of S. The payoff of players A is exactly the sum asked for themselves (a1, a2 and a3 respectively), while the payoff of player B is the remaining sum ($S - a1 - a2 - a3$). In our experiment, each group of four players has 60 tokens to allocate – each token corresponds to 50 eurocents.

The experiment consists of three treatments: the *Baseline Treatment (BT)*, the *Agreement Treatment (AT)* and the *Outsider Treatment (OT)*.

In the *BT* participants are randomly distributed in groups of four players and play the *Exclusion Game*.

In the *AT* participants are randomly distributed in groups of four players and are instructed about the stages of the experiment and about the *Exclusion Game*. In the first stage, before knowing their role in the game they are

¹ We have Empirical Expectations (EE) when a subject believes that a sufficiently large subset of the population conforms to the norm. We have Normative Expectations (NE) when a subject believes that a sufficiently large subset of the population expects him/her to have to conform to the norm.

involved in a voting procedure. In particular, in each group participants are invited to vote for a specific allocation rule². Players have the possibility to bargain - through the computer - to reach a unanimous agreement on the rule within a limited numbers of trials (10 in our experiment). The agreement is not binding, but failure in reaching it is costly, since only groups who reach an agreement in this first stage have the chance to participate to the second stage. In the second stage the composition of the groups is unchanged and roles are randomly assigned to implement the *Exclusion Game*. In this case, players A can decide either to implement the voted rule or to choose one of the alternative allocations. Players who do not enter the second stage wait for the end of the session. Their payoff is the show-up fee.

In the *OT* participants are randomly distributed in groups of four players and are instructed about the stages of the experiment and about the *Exclusion Game*. The first stage as well as the rule to enter the second stage are the same as in the *AT*. At the beginning of the second stage, players are informed about their role and groups are rematched. In particular, a player A for each group (the outsider) is reassigned to a different group and instructed about the rule chosen by the new group. After the rematching subjects participate in the *Exclusion Game*. Also in this case players who do not enter the second stage wait for the end of the session and they are paid only the show-up fee.

For a summary see Table 1.

Experimental Procedure. The experiment was run both in Milan (EELAB – University of Milan Bicocca) and in Trento. Recruitment was done by email advertisement. We ran 3 sessions for the *BT* (1 in Milan and 2 in Trento), 4 sessions for the *AT* (2 in Milan and 2 in Trento), 5 sessions for the *OT* (3 in Milan and 2 in Trento). Overall, 216 undergraduate students – 104 in Milan and 112 in Trento – participated to the experiment. A more detailed description of the sessions is in Table 1.

² Subjects must vote one out of three alternative division rules (the forth number is player B's payoff): {15,15, 15,15},{18,18, 18,6}, {20,20, 20,0}. The first rule assigns the same payoff to every member of the group; the second rule corresponds to a partial inclusion of player B in sharing the wealth; the third rule implies the total exclusion of player B.

The experiment was programmed and conducted with the software z-Tree (Fischbacher, 2007). The instructions were read by participants on their computer screen while an experimenter read them loudly.

After reading the instructions and before subjects were invited to take decisions, some control questions were asked in order to be sure that players understood the rules of the game. At the end of each session, subjects were asked to fill in a brief survey to check for socio – demographic data.

Players were given a show – up fee of 3 euro.

Beliefs elicitation. In all the treatments, at the end of the game and before players are informed about the decisions taken during the *Exclusion Game* by the other coplayers, first order and second order expectations (both normative and empirical) are elicited through a brief questionnaire. In particular, in each group each player makes a statement: 1. of the probabilities related to each possible choice of coplayers A (First Order Empirical Expectations); 2. of the probability related to each coplayers' possible judgement about his/her own choice (Second Order Empirical Expectations); 3. of the choice should have been taken by a representative player A (First Order Normative Expectations) ; 4. of the choice that coplayers consider as the 'right' one (Second Order Normative Expectations).

Both in the *AT* and in the *OT* only players who enter the second stage are interviewed about their expectations. Moreover, in the *OT* guesses on behaviour and beliefs of partners and outsiders are asked separately.

Only good guesses of the Empirical Expectations are rewarded through the well-known quadratic scoring rule.

Results.

Result 1. Subjects' choices are in line with their expectations.

If we check whether there is any correlation between beliefs and decisions, it turns out that most players' choices are in line with either empirical or normative expectations (Table 2).³

However - as in Bicchieri and Xiao (2007) - when normative and empirical expectations are in contrast, the latter play a more relevant role in players' decisional process (Table 3) and they are significantly correlated to subjects' choices (Spearman test; $p < 0.03$).

Result 2. When agreement is possible, it is reached by all groups.

See Table 4 for a detailed description of agreements.

Result 3. Agreement induces convergence of empirical expectations.

In the BT at least 70% of the players choose 20. In the AT 17 groups out of 18 choose the 15-15-15-15 rule and 1 the 18-18-18-6 one. If we analyse people's expectations, it turns out that in the AT there is a significant decrease of subjects who think that the other members of their group have asked 20 tokens (Table 5). A probit regression shows that the probability of expecting the others have chosen 20 is significantly lower in the AT ($p = 0.000$). Consequently, also choices are significantly lower in the AT (Mann-Whitney; $p = 0.0002$).

Result 4. Expectation of conformity is higher in the partner protocol.

If we analyse the difference between players' empirical expectations about other players' choice and the amount to be chosen according to the voted rule, it turns out that it is significantly lower in the AT (OLS; $p = 0.054$). This means that in the OT subjects expect a higher deviation from the chosen rule. Since expectations and choices are correlated, this explains why a lower percentage of players comply to the norm in the OT (Table 6).

³ We consider only first order expectations since second order expectations are either equal or highly correlated to the former.

References

- Bicchieri, C. (2006), *The Grammar of Society: The Nature and Dynamics of Social Norms*, Forthcoming Cambridge University Press.
- Bicchieri, C., Xiao E., (2007), *Do the Right Thing: But only if others do so*, mimeo.
- Bochet O., Page T., Putterman L. (2005), Communication and Punishment in Voluntary Contribution Experiments, Working Papers 2005-09, Brown University, Department of Economics.
- Camerer, C.F. (2003), *Behavioral Game Theory: Experiments in Strategic Interaction*, Princeton; NJ; Princeton University Press.
- Dufwenberg M., Gächter S., Hennig – Schmidt H. (2006), The Framing of Games and the Psychology of Play, Discussion Papers 2006-20, The Centre for Decision Research and Experimental Economics, School of Economics, University of Nottingham.
- Faillo, M. and Sacconi L. (2007), “Norm Compliance: The Contribution of Behavioral Economics Theories”, *Discussion Paper. Department of Economics University of Trento*. Forthcoming in Innocenti, A. and Sbriglia, P. *Games, Rationality and Behavior*, Palgrave.
- Fehr E., Fischbacher U. (2004), ‘Third Party Punishment and Social Norms’, *Evolution and Human Behavior*, 25, 63-87.
- Fehr E., Gächter S. (2000), ‘Cooperation and Punishment in Public Good Experiments’, *American Economic Review*, 90(4), 980-994.

- Fehr, E. and K.M. Schmidt (2000) "Theories of Fairness and Reciprocity. Evidence and Economic Applications", *Institute for Empirical Research in Economics University of Zurich Working Paper Series*, n.75.
- Geanakoplos, J., D. Pearce and E. Stacchetti (1989), "Psychological Games and Sequential Rationality", *Games and Economic Behavior*, Vol. 1, pp. 60-79.
- Grimalda, G., L. Sacconi (2005) "The Constitution of the Not-for-Profit Organisation: Reciprocal Conformity to Morality", *Constitutional Political Economy*, Vol 16(3), 249-276.
- Güth W., Levati M.V., Ockenfels A., Weiland T. (2005), "'Buying a pig in a poke": An experimental study of unconditional veto power," Discussion Papers on Strategic Interaction 2005-39, Max Planck Institute of Economics, Strategic Interaction Group.
- Rabin M., 1993, 'Incorporating Fairness into Game Theory and Economics', *American Economic Review*, 83(5), 1281-1302.
- Sacconi, L. and Faillo, M. (2005), "Conformity and Reciprocity in the "Exclusion Game": An Experimental Investigation" *Discussion Paper. Department of Economics University of Trento*.
- Sacconi L. and G. Grimalda (2007) "Ideals, conformism and reciprocity: A model of Individual Choice with Conformist Motivations, and an Application to the Not-for-Profit Case" in (L.Bruni and P.L.Porta eds.) *Handbook of Happiness in Economics*, Edward Elgar, London (in print).

Table 1. Experimental Design

Treatment	Voting Procedure	Matching	Sessions	Subjects
BT	NO	Partner Protocol	2 in Trento (T) 1 in Milan (M)	36 (T) + 20 (M) 9 groups (T) + 5 groups (M) (27 (T) + 15 (M) players A)
AT	YES	Partner Protocol	2 in Trento (T) 2 in Milan (M)	36 (T) + 36 (M) 9 groups (T) + 9 groups (M) (27 (T) + 27 (M) players A)
OT	YES	Mixed – Partner and Stranger Protocol	2 in Trento (T) 3 in Milan (M)	32 (T) + 56 (M) 8 groups (T) + 14 groups (M) (24 (T) + 42 (M) players A)

Table 2. Beliefs and choices

	It is possible to explain subjects' behaviour through...		
	FOEE	FONE	OTHER
BT			
T (N = 27)	82%	7%	11%
M (N = 15)	93%	0%	7%
AT			
T (N = 27)	82%	11%	7%
M (N = 27)	82%	7%	11%
OT			
T (N = 24)	71%	21%	8%
M (N = 42)	83%	10%	7%

Table 3. Normative and empirical expectations

When FOEE and FONE are different it is possible to explain subjects' behaviour through...			
	FOEE	FONE	OTHER
BT T (N = 14) M (N = 8)	72% 100%	14% 0%	14% 0%
AT T (N = 11) M (N = 9)	64% 78%	27% 22%	9% 0%
OT T (N = 14) M (N = 21)	57% 71%	14% 19%	29% 10%

Table 4. Group's Choices by University x Treatment.

		Rule					
		15 – 15 – 15 – 15		18 – 18 – 18 – 6		20 – 20 – 20 – 0	
Trento	AT	88.9%	8/9	11.1%	1/9	0.0%	0/9
	OT	87.5%	7/8	12.5%	1/8	0.0%	0/8
Milano	AT	100.0%	9/9	0.0%	0/9	0.0%	0/9
	OT	92.9%	13/14	0.0%	0/14	7.1%	1/14

Table 5. Distribution of FOEE by University x Treatment

		15 - 18	20
Trento	BT (N = 27)	15.0%	85.0%
	AT (N = 27)	20.0%	80.0%
Milano	BT (N = 15)	52.0%	48.0%
	AT (N = 27)	69.0%	31.0%

Table 6. Compliance by University x Treatment.

Trento	AT	44.4%	12/27 10 rule 15 - 2 rule 18
	OT	29.2%	7/24
	<i>OT</i> (Insiders)	37.5%	6/16 5 rule 15 - 1 rule 18
	<i>OT</i> (Outsiders)	12.5%	1/8 1 rule 15
Milano	AT	55.5%	15/27 15 rule 15
	OT	45.2%	19/42
	<i>OT</i> (Insiders)	39.3%	11/28 9 rule 15 - 2 rule 20
	<i>OT</i> (Outsiders)	57.1%	8/14 7 rule 15 - 1 rule 20

“When in Rome”: Applying a method for eliciting social norms

Erin Krupka, Roberto Weber and Rachel Croson

Over the last half decade economists have produced compelling research which highlights the role that social norms can play in explaining behavior (Akerlof 1980; Lindbeck et al. 1999; 2006). The increased attention to social norms in economics has also generated important questions regarding what social norms are and how they might be (systematically) and empirically identified across domains and choice contexts.

In this paper, we explore the critical role that jointly held beliefs play in defining and eliciting social norms. In particular, we explore whether we can use an instrument, which relies on coordination games, to systematically identify social norms. Building on previous research, we define a social norm as a socially shared agreement regarding the appropriateness or inappropriateness of a behavior. This definition has wide support in the social science literature but gives particular prominence to the notion that social norms reflect *shared beliefs* rather than personal beliefs (Elster 1989; Bettenhausen and Murnighan 1991; Fehr and Schmidt 1999; Lindbeck et al. 1999; Bicchieri 2000; Bicchieri and Chavez 2007). This, in turn, suggests that social norms can be identified by asking for shared beliefs rather than personal beliefs. Second, this definition implies that a reference social group critically determines which beliefs are jointly held. And thirdly, it suggests that we might expect jointly held beliefs to change when the reference social group changes (Elster 1990; Margolis 1996; Miller and Prentice 1996). These three observations yield three predictions which we test.

- 1) Reference group hypothesis: When social norms vary between two reference groups, then an individual will judge the same behaviors to be differently appropriate depending on which reference group he uses.
- 2) Norm tracking hypothesis: If an individual's appropriateness judgments differ depending on which reference group is used, then their stated judgments should covary with the actual, independently and ex-ante identified social norms associated with each group.
- 3) Socially shared agreement hypothesis: appropriateness judgments obtained without the use of a coordination game capture personal beliefs. Personal judgments of appropriateness will not significantly explain the variation in appropriateness judgments obtained when reference groups change.

To test these predictions, we directly elicit shared beliefs using coordination games. We present US born and non-US born subjects with two scenarios involving an hypothetical actor “A”, who must decide what percentage of a \$10 lunch bill to tip (a tipping scenario)

and whether to arrive early, on time, or late to a meeting at the library (a punctuality scenario). In each scenario, we describe seven different actions that “A” could take and instruct subjects to judge and rate the social appropriateness of each of those seven actions.

Participants first read about an “individual A” in a punctuality scenario, and then they are asked to rate the appropriateness of each of A’s actions while (1) matching their ratings with those of one other (randomly selected) campus student from the same university and then (2) while matching their ratings with those of one other (randomly selected) student who was born in the same country as the subject. Participants were informed that one survey would be selected at random and that, if their survey was the one selected, then the experimenters would pay them an additional \$1 for every response that was *exactly the same* as that of the person which whom they are trying to match. The selected participant would also receive an additional \$20 if the entire survey was completed. This means that the selected subject could earn the \$10 show up fee that everyone received and up to another \$48 if his survey was selected (\$20 for filling out the entire questionnaire and \$28 if all responses matched).

After completing the two matching tasks for the first scenario, the subject also rated each of A’s actions according to their own personal beliefs. After completing both matching tasks and stating their own personal beliefs for the punctuality scenario, the entire sequence was repeated for the tipping scenario. The order of scenarios was never changed. That is, subjects always saw the punctuality scenario first and the tipping scenario second.

Table 1: Overview of the Experimental Design

	Matching ratings with:		
	Another student on campus	Another student born in the same country	State own beliefs
Punctuality Scenario			
Tipping Scenario			
	\$ Incentives used	\$ Incentives used	No incentives
	Order counter-balanced		

Thus, subjects make a total of six sets of appropriateness ratings. However, the order in which subjects were asked to match with another campus student or with a fellow native student was counterbalanced for half the subjects (see Table 1). All subjects were always asked to record their personal beliefs last (column 4 in Table 1) after having completed the matching tasks for each scenario.

In the punctuality scenario subjects read the following short description:

Individuals A and B have agreed to meet in the library to work on an assignment. They have agreed to meet at a particular time. Individual A has to decide at what time to arrive for the meeting. Individual A has seven possible choices which are indicated in the table below. Individual A can choose only one of these options.

Individual A's possible actions were then listed in a table and the subjects were asked to rate each action on an appropriateness scale while trying to coordinate their ratings with either another campus student or a student born in their native country. That is, subjects rated how appropriate it is for A to arrive 10 minutes early, 5 minutes early, exactly on time, 5 minutes late, 10 minutes late, 20 minutes late or 30 minutes late. Table 2 is a portion taken from the tables subjects used to record their appropriateness ratings for the punctuality scenario. The two lines on the table are exactly as our raters would have seen them.

Table 2: Example of the ratings form used for the punctuality scenario

Individual A's choice	Very socially inappropriate	Somewhat socially inappropriate	Somewhat socially appropriate	Very socially appropriate	(Experimenter use only)
					Response matches partner's?
Arrive exactly on time.					Y N
Arrive 20 minutes late.					Y N
					Total Matches: Additional \$\$ Earned:

In the tipping scenario each subject read the following short description:

Individual A has just finished lunch at a restaurant. The bill for lunch is \$10.00. Individual A must decide how much, if any, tip to leave. Individual A has seven possible choices which are indicated in the table below. Individual A can choose only one of these options.

Individual A's actions are to give no tip (A pays \$10 and a \$0 tip), give a 5% tip (A pays \$10 and a \$0.50 tip), give a 10% tip (A pays \$10 and a \$1 tip), give 12% tip (A pays \$10 and a \$1.20 tip), give 15% tip (A pays \$10 and a \$1.50) tip, give 20% tip (A pays \$10 and a \$2.00 tip).

Because both tipping and punctuality norms are very common, we anticipate that subjects are likely to be familiar with them. Further, because we have a very good ex-ante idea of the expected (actual) real-world social norm for each scenario we are able to directly assess how well the experimentally elicited appropriateness ratings track true tipping and punctuality norms. We exploit the fact that, for the *same* hypothetical scenario, tipping and punctuality norms will vary from country to country. As a result, we observe how ratings *for the same subject* change depending on whether that subject is matching his/her

appropriateness ratings with a campus student from the same university or with another student who was born in the same country as the subject. In addition to asking subjects to match their responses to different reference groups, we ask subjects to rate how appropriate they *personally* believe each action is so that we may compare appropriateness ratings elicited using a matching task to ratings elicited by asking about personal beliefs.

Consistent with our 'Reference group hypothesis', we show that subjects' appropriateness ratings depend on whether they are instructed to match their ratings with another campus student or another student who was born in their native country. We also find that the elicited appropriateness ratings vary in a manner consistent with the actual country-specific social norms associated with tipping and punctuality ('Norm tracking hypothesis'). Finally, we demonstrate that simply asking subjects to state their personal beliefs produces ratings that differ significantly from those elicited using the norm elicitation technique ('Socially shared agreement hypothesis'). We suggest that this is because the matching procedure focuses subjects on the social component of the appropriateness of each behavior and thus elicits a social norm rather than a personal norm. We take these three results together to argue that the elicitation instrument used in the experiment accurately identifies social norms.

Extended abstract

Context

Despite of the important efforts and improvements against the HIV/AIDS epidemic made by scientists, political actors, and citizens, the number of people who is living with the HIV is still increasing, since in 2006 there was 2.6 millions of infected people more than in 2004 all over the world (UNAIDS, 2006). The epidemic is far from being under control, especially in sub-Saharan Africa, East Europe, and Central Asia.

The relevance of prevention is even greater when we take into account that no curative treatment has been created. For that reason, many authors have investigated about the factors that influence the adoption of preventive practices, such as the decrease in sexual partners, sexual abstention, and especially the condom use, in different societies (Caldwell, 1999; Eaton et al., 2003; Stoneburner y Low Beer, 2004). A paradoxical fact that has been observed is that individuals who are aware of the risk associated with HIV/AIDS and who have a relatively good knowledge about the disease, the ways the infection can be transmitted, and the range of preventive practices that can be adopted, they do not necessarily protect themselves during their sexual relationships. Many authors have pointed out that this question could be understood if the relevant role of social factors in the decision of, for instance, using a condom is taken into account (Dolcini, et al., 2004; Gausset at al., 2001; Rushing, 1995).

Several theoretical and methodological frameworks have been used in the research about the effect of social factors on risk (preventive) sexual behaviour. At first, studies were predominantly based on cultural perspectives that tended to blame, especially in the analysis of Africa, certain cultural characteristics, the conception of death, and the widespread promiscuity for the levels of prevalence. Afterwards, a great amount of studies has taken a social psychology approach, so that they have emphasized the importance of interpersonal communication about sexuality and AIDS in learning about prevention and in the development of communicative skills, which facilitate the negotiation process of protected sex. Most of these researches have used qualitative techniques in the study of the influence of communication with the partner on the use of condom among adolescents and ethnic minorities in the US (Faulkner, 2002; Gómez et al., 1999; Talashek et al., 2003, Whitaker at al, 1999). Some authors have tried to understand the role of social interactions from sociological and demographical perspectives, but it is not easy to find rigorous quantitative analyses of the mechanisms through which interpersonal communication affect individual's decision about protected sex.

Objective of the research

The theoretical argument behind this analysis is that the spread of a new sexual practice such as the condom use in a society can be understood as a new social norm, which can conflict with established social norms about sexuality, gender relationships, or even illness. This makes sense when we conceive this question as a coordination problem, according to which people's behaviour is influenced by shared expectations as regards what should/should not be done in specific situations. These expectations depend, in turn, on individuals' beliefs about the proportion of the population of reference who has already adopted this new practice, the condom use, in different kinds of sexual relationships. Finally, individuals' beliefs and expectations are constantly updated in their social interactions. That is why the study of verbal communication becomes crucial in the understanding of the problem. Open communication with other people about AIDS, sexuality or related subjects may facilitate the spread process of a new preventive practice, especially in social contexts where the condom use conflicts with certain social norms already established. For that reason, this paper is an attempt to go more deeply into the analysis of the influence of social interactions in the use of condom, considering aspects such as the types of confidants and their links with the individual, as well as the level of knowledge about the disease and attitudes towards AIDS among the people around her. Besides, the theoretical argument adds that a new social climate, derived from open communication and characterized by more positive attitudes towards preventive sexual practices against AIDS among the community, may encourage the change in sexual behaviour through favouring and making easier the communication about the question with the sexual partner. That is why this aspect is taken into account in the analysis as an intermediate variable. In my opinion, this empirical analysis may shed some light on the unsolved question about the specific mechanisms through which social norms affect individuals' decisions.

Methodology and data

Unlike many researches about communication, the empirical research is based on the use of quantitative techniques, in order to take advantage of the valuable information in the Demographic and Health Surveys (DHS). The most appealing feature of the DHS is that they have been conducted in many different developing countries, so they allow to make a comparative research of women between 15 and 49 years old in diverse social, political and economic contexts. Moreover, in most of these countries, the survey has been carried out in

more than one moment on time, so the observation of changes in behaviour and attitudes within the same society is possible. My intention is to focus on two of the countries with the highest levels of HIV prevalence in sub-Saharan Africa, which have lived very different experiences with the epidemic: Uganda and Malawi.

Bibliography

- Caldwell, John C.** 1999. "Reasons for Limited Sexual Behavioural Change in the Sub-Saharan African AIDS Epidemic, and Possible Future Intervention Strategies." Pp. 241-56 in *Resistances to Behavioural Change to Reduce HIV/AIDS Infection*, edited by John C. Caldwell et al. Canberra: Health Transition Centre.
- Dolcini, Margaret, Lisa Canin, Alice Gandelman, and Heidi Skolnik.** 2004. "Theoretical Domains: A Heuristic for Teaching Behavioral Theory in HIV/STD Prevention Courses." *Health Promotion Practice* 5:404-417.
- Eaton, Liberty, Alan J Fisher, and Leif E Aaro.** 2003. "Unsafe Sexual Behaviour in South African Youth." *Social Science and Medicine* 56:149-65.
- Faulkner, S. L.** 2002. "Reconciling Messages: The Process of Sexual Talk for Latinas." Gausset, Quentin. 2001. "AIDS and Cultural Practices in Africa: The Case of the Tonga (Zambia)." *Social Science and Medicine* 52:509-18.
- Gómez, C. A., M. Hernández, and B. Faigles.** 1999. "Sex in the New World: An Empowerment Model for HIV Prevention in Latina Immigrant Women." *Health Education & Behavior* 26:200-212.
- Qualitative Health Research* 12:310-328.
- Rushing, William A.** 1995. *The AIDS Epidemic: Social Dimensions of an Infectious Disease*. Boulder (Col): Westview Press.
- Stoneburner, RL , and D. Low-Beer.** 2004. "Population-level HIV Declines and Behavioral Risk Avoidance in Uganda." *Science* 304:714-18.
- Talashak, M. L., K.F. Norr, and B.L. Dancy.** 2003. "Building Teen Power for Sexual Health." *Journal of Transcultural Nursing* 14:207-216.
- Whitaker, D. J., K.S. Miller, D.C. May, and M.L. Levin.** 1999. "Teenage Partners' Communication about Sexual Risk and Condom Use: The Importance of Parent-Teenager Discussions." *Family Planning Perspectives* 31:117-121.

Communication, Trust and Reciprocity among Members of Ethnic Groups in Conflict

A Paper Proposal for the XI Summer School on Economics and Philosophy, San Sebastian (Spain), 14-17 July 2008.

Dr. Azi Lev-On

Head of New Media Track, School of Communication, Ariel University Center,
Israel

My long-term interest in the social and political implications of Internet communication, and the life in an area fraught with conflicting ethnic groups, have lead me to pursue an experimental project dedicated to researching the various potential possibilities for generating trust, reciprocation and cooperation through communication using new media, with specific regards to exchanges among Jews and Arabs.

In circumstances of deep hostility and distrust between populations, especially where the barriers to exchange are not only emotional but geographical and physical as well, computer-mediated communication (specifically the Internet) enables Jews and Arabs to more easily gain exposure to each others' cultures, beliefs, opinions and daily lives. A few studies have examined what happens when Jews and Arabs can deliberate or even just talk amongst each other regardless of the topic (i.e., not limited to politics.) The results of such interactions are generally encouraging in terms of knowledge, tolerance and willingness to continue and explore the unique characteristics of the other side (e.g. Ellis and Maoz 2007, Yablon and Katz 2001).

But participation in such discursive settings typically entails high cognitive and opportunity costs from participants (and hence may require a large monetary investment from the organizing institution if carried out outside the experimental lab and utilized on a large scale.) Alongside such 'expensive' settings, it is quite likely that in time, exchanges between members of rival groups will involve many more singular interactions between members of the two populations. However, in these settings, only a few cues about the identity of one's interlocutor are available (for example one's email address, place of living or preferred language) as is done on e-commerce sites. These are the settings that I study.

In spite of the relative 'poverty' of such exchange environments in comparison to the 'richer' and 'thicker' deliberative online arenas, they still introduce a variety of potentially beneficial effects. This holds in the macro-level (the classic story being Adam Smith's theory about the civilizing effects of commerce, and its contemporary successors), and on the micro-level as well. Notably, the contact hypothesis – which is one of the leading theories on managing inter-group conflicts - suggests that interactive and direct communication between members of rival groups will lead to the reduction of prejudices (Amichai-Hamburger and McKenna 2006). The question is, under what circumstance individuals will be able

to build trust online and enjoy the mutually beneficial social and individual advantages of exchange (Lev-on 2007).

To study the emergence of trust in such exchange scenarios, I will employ experimental 'trust games' similar to the ones in Bicchieri, Lev-On and Chavez (2007) and in Lev-On, Chavez and Bicchieri (2007). The experimental laboratory is a first-rate environment to study trust and reciprocation as it allows for the manipulation and control of multiple variables, the separation of their effects and the testing of competing hypotheses. In the proposed experiments subjects are assigned one of two roles: first-movers and second-movers. The experiments include two decision periods. First, each first-mover receives an endowment and decides to send some, all, or none of it to the second-mover. The amount the first-mover does not send is theirs to keep, while the amount sent is multiplied by the experimenters. Second, the second-mover decides to send some, all, or none of this amount to the first-mover. The amount the second-mover does not send is hers to keep. If trust is established and honored, both players receive a dividend on their 'investment', which is not generated at the absence of trust. *Let me note that the games are played for real money and subjects leave the lab with the actual amount they earn.* Such experiments are used frequently in the social sciences to study issues related to trust and reciprocation.

A robust experimental finding in lab experiments (in which subjects are typically students who come from similar backgrounds) is the facilitative effect of communication on cooperation, which Bicchieri and Lev-On (2007) denote as the 'communication effect'. In such experiments, computer-mediated communication and even more so face-to-face communication, allow subjects to make mutual promises which are perceived as reliable and set the cognitive groundwork for cooperation. Consequently, cooperation rates following communication are much higher than absent communication, even though there are no mechanisms for enforcing promises. Subjects know that communication can be used strategically to convince them to deliver funds and then betray their trust.

The proposed experiment is slightly different from the typical experiments described above as subjects belong to two distinct ethnic groups: Jews and Arabs. Subjects will be recruited through clicking on a banner placed in the homepage of the academic institute's student union, through direct mail correspondence, and through flyers that will be distributed outside the building where the experiments take place, in the day of the experiment.

In the laboratory, participants will have read and received written general instructions about the trust game. Then they will be divided into pairs, and (in the communication conditions) engage in communication with the person with whom they are paired. In the face-to-face conditions, the communication period will be two minutes long. In the CMC conditions, subjects will have five minutes to communicate via a computer chat program. In some experimental sessions the pair of subjects will be composed of a Jewish subject and an Arab subject (Jewish * Arab – i.e. the Jewish subject is the first-mover, the Arab subject is the second-mover). In other sessions the pairing is: Arab * Jewish, Arab * Arab, and Jewish * Jewish. The experimental design is thus $2*2*2$ (ethnicity of first-mover * ethnicity of second-mover * communication medium). Note again that the game is played

for real money; thus, one's decision to trust or reciprocate has immediate monetary implications for the welfare of the subjects with whom he/she is paired.

Note that the experiment has obvious implications for the possibilities to generate trust and reciprocity between members of rival groups through advanced communication means. It is interesting to compare our results with results of 'standard' studies involving communication, where subjects share a common background, and to learn if computer-mediated communication can build trust between 'different' subjects, or just support cooperation between 'similar'.

References

- Amichai-Hamburger, Y., & McKenna, K. Y. A. (2006). The contact hypothesis reconsidered: Interacting via the Internet. *Journal of Computer-Mediated Communication* 11(3), article 7. <http://jcmc.indiana.edu/vol11/issue3/amichai-hamburger.html>
- Bicchieri, C., and Lev-On, A. (2007). Computer-mediated communication and cooperation in social dilemmas: An experimental analysis. *Politics, Philosophy and Economics* 6: 139-168.
- Bicchieri, C., Lev-On, A., and Chavez, A. (2007). The medium or the message? Communication richness and relevance in trust games. Submitted for publication.
- Ellis, D., and Maoz, I. (2007). Online Argument between Israeli Jews and Palestinians. *Human Communication Research* 33: 291-309.
- Lev-On, A. (2007). Cooperation with and without trust online. Forthcoming in *Trust and Reputation*, ed. Karen Cook. Available at <http://homepages.nyu.edu/~el322/downloads/Lev-On%20Coop%20withuot%20Trust%20Online.pdf>
- Lev-On, A., Chavez, A., and Bicchieri, C. (2007). Group and dyadic communication in trust games. Submitted for publication.
- The Good Neighbors Blog , www.gnblog.com
- Yablon, Y.B., and Katz, Y. J. (2001). Internet-Based Group relations: A High School Peace education Project in Israel. *Education Media International*. 38(2/3): 175-182.

**Social Connections, Group behaviours and Corruption:
An Experimental Study of a New 'Favour Game'**

Donna Harris

Environmental Economy and Policy Research Group
Department of Land Economy
University of Cambridge

January 4, 2008

Despite extensive research on the causes and consequences of corruption, most studies have not systematically addressed the antecedents of corruption from a behavioural perspective - that is 'why do people decide to engage in a corrupt transaction?' Corruption is generally defined in the literature as 'the abuse or misuse of public power for private gain' (Jain, 2001). However, taken the specific context and society into account, the exact meaning of corruption is open to discussion and depends upon social constructions. In practice, corruption can take various forms. The focus of this study is on the type of corruption which relies heavily on personal connections and existing social networks in which social identity plays a central role, namely 'nepotism'. Nepotism is defined in this study as the in-group favouritism behaviour which is not only unfair, but also imposes negative externalities on the out-group. This type of corruption is prevalent both in the public and private sectors and can lead to inefficiency due to misallocation of resources, particularly within developing countries. However, it has received relatively less attention in the literature and has not been tested empirically.

This study will use a newly designed economic experiment to shed a light on the relationship between social network and nepotism behaviour by considering the mechanism by which social connections and the variant in the strength of such connections shape group behaviours (in-group vs. out-group), and how such behaviours influence the individual's decision whether to engage in nepotism behaviour. Specifically, I will address the following questions:

- To what extent do people from different cultural backgrounds perceive a similar range of activities as corrupt or non-corrupt and how do such perceptions affect their nepotism behaviour
- To what extent does the strength of social connections within a group affect nepotism behaviour?
- Does the threat from the *third-party* or altruistic punishment influence nepotism behaviour and if so, to what extent does the strength of social connection within a group affects this mechanism?
- Does the threat from the *out-group* punishment (those who directly suffer) influence nepotism behaviour and if so, to what extent does the strength of social connection within a group affects this mechanism?
- Does the threat from *in-group* punishment affect nepotism behaviour and if so, to what extent does the strength of social connection within a group affects this mechanism?

The main objective of the game is to firstly examine the relationship between variant in the strength of social connections within group and nepotism and secondly, to investigate the effects of the threats from different types of punishment, namely altruistic punishment from the third party who do not suffer directly from nepotism, punishment from those who lose out as a result of this form of corruption, and of central interest to this study, the threat of punishment from the in-group on its members to the individual who has to decide whether to engage in nepotism (the public official, for example) if they do not favour the group (engaging in nepotism). The main conjecture is that the stronger the level of social connection within the in-group, the higher the propensity of nepotism behaviour, despite the damage it may cause to the others who do not belong to the in-group. This proposition is likely to hold despite the threats from the out-group and altruistic punishments. This is because the stronger the social connections among the in-group members, the stronger the alternative set of social norms that highly value particularised reciprocity and loyalty to other members of the group are enforced. The cost of deviating from such in-group norms is likely to be large and the in-group punishment is severe to the extent that it may outweigh the costs of punishments from the outsiders and the third-party. This in-group punishment generally takes the form of severe social sanction such as ostracism, which creates disutility for the member who deviates from the in-group norms. In addition, once a member of the corrupt network, it is very difficult to exit. The inside information that the exiting member has obtained about the corrupt network means that other members would do everything in their power to prevent such member from exposing them as corrupt.

Urrutia Elejalde Foundation

Summer School in Economics and Philosophy

TEAM FORMATION EXPERIMENTS AND THE FAIRNESS NORM AN INTERNATIONAL COMPARISON

Brice Corgnet
Universidad
de Navarra

Angela Sutan
Burgundy
Business School

Robert Veszteg
Universidad
del País Vasco

Extended abstract

I. The fairness norm in bargaining: equity versus equality

In this paper, we analyze team formation in a real effort experiment in which individuals bargain their share of the team outcome. We deliberately use a real task in order to induce effort and merit in our analysis. In most economic experiments subjects are involved in abstract settings in which exerting an effort consists of losing a certain amount of money. We aim at understanding in the current study the difference between the social norms of *equity* and *equality*. Both terms are usually confounded under the elusive concept of *fairness* but *equality* and *equity* are fundamentally different concepts once we take into consideration the issue of effort and merit. The *equality* or *egalitarian* norm implies that partners should be rewarded an equal share of the joint outcome. The equal split of the team output is frequently observed in organizations, partnerships, joint ventures and share tenancy in agriculture and constitutes a puzzle for standard economic theory (Andreoni and Bernheim 2007). The *equity* norm implies that partners should be rewarded based on their relative contributions. In that case, team members are paid according to their merit that is typically measured by their level of effort and their ability. This distinction between *equity* and *equality* raises the following question. Should we consider that sharing a joint outcome equally among the different contributors as *fair*? In most of the experiments in the Economics literature like the ultimatum game (Güth et al. 1982), the joint outcome is not the result of a real effort but simply an amount of money provided by the experimenter at the beginning of the session. In that context, *equity* and *equality* would both imply an equal sharing of the initial amount of money. In a case in which subjects did not contribute to the joint outcome there is no discussion about the relative merit of each partner. In the ultimatum game, a subject (the *proposer*) offers a division of a fixed amount of money received from the experimenter between herself and a *responder*. The *responder* can reject the offer, in which case neither subject earns

anything. If the *responder* accepts the division of the initial amount of money then it is implemented. In these experiments, close to half of the *proposers* offer an equal split of the team outcome. Also, responders often reject unequal divisions of the initial amount of money. These results are interpreted as evidence of a *fairness* norm in bargaining (Fehr and Schmidt 1999), where *fairness* coincides with the *equality* norm in that case. Interestingly, in the context of ultimatum games, the *equality* norm is found to be robust to Japanese subjects (Roth et al. 1991, Oosterbeek et al. 2004) and to a number of small scale societies (Henrich et al 2004). However, in real life situations individuals have different abilities and exert different levels of effort that may lead to very different outcomes. As a result, we propose an experiment in which individuals exert a real effort that may have different consequences on the joint profits according to the levels of ability and effort of the team members. This experimental design allows us to separate the norms of *equity* and *equality*. Our purpose relates to the analysis of Schurter and Wilson (2007) that consists in separating the effects of *fairness* and *justice* by analyzing the role of status on the behavior of subjects in a dictator game. In this context, the authors relate *justice* to merit (*equity*) while emphasizing that *fair* procedures may not be based on merit as it is the case if they lead to the systematic use of the *equality* norm. The authors find that merit is a determinant variable in the patterns of offers by dictators as it significantly reduced the average offer. The main difference between our notion of *equity* and the concept of *justice* is that the norm of *equity* does not need an external authority such as a court to be implemented. *Equity* is interpreted as an implicit norm that may arise in the absence of legal enforcement.

In our setting, the subjects undertake a real task and are then asked to assess the relative contribution of each individual. We are able to evaluate the impact of merit on the choice of the sharing rule of the joint outcome by assessing partners' relative contributions.

II. Culture and the *fairness* norm

The Business and Psychology literature stresses important cultural differences in the context of *reward allocation experiments* (Fischer and Smith 2003). These virtual experiments describe scenarios in which a number of individuals work on a task and exhibit different levels of performance. Subjects are then asked about an allocation of the rewards to the different individuals. As it is stressed by James (1993), in that context subjects behave differently according to certain cultural dimensions such as individualism and collectivism. In particular, subjects from individualist cultures tend to follow the *equity* norm whereas subjects pertaining to collectivists cultures use the *equality* norm extensively. However, these studies have been criticized on their internal validity and in particular on the fact that many cultural dimensions concurrently affect social norms and then reward allocation. Our experimental analysis constitutes an alternative and independent approach that may shed light on the robustness of international differences on social norms of *equity* and *equality*. Besides the methodological aspect, a crucial difference with our approach is the fact that in these studies reward allocation is decided by an external observer that was not affected by the result of the allocation. This situation typically represents a monitoring environment whereas our experimental design illustrates the case of team partners that bargain over the allocation of a joint outcome that affects each individual's payoff.

III. Experimental design

We design an experiment in which subjects undertake a real task for which they may possess different levels of ability. The task consists in finding numbers that satisfy certain restrictions. Every correct answer is rewarded in cash and the total amount of money accumulated during the task constitutes the benefits of the group. Subjects first worked in pairs that were formed randomly at the beginning of the experiment. In this first task, team partners shared their profits equally. After the subjects had been isolated, they were asked a series of questions. They first had to assess the perception of their own contribution to the group outcome. Then, they had to state the minimum share of the joint outcome that they would accept in order to work with the same partner in the next task. The second task was performed either individually or by pairs depending on the compatibility of group members' claims. We considered that the claims of two partners were compatible if both subjects were ready to accept a sharing rule of the joint outcome randomly taken from a bowl. The bowl contained nine pieces of paper that represent the following sharing rules: 20%, 30%, 40%, 45%, 50%, 55%, 60%, 70%, and 80%, where each number indicates the share of the joint outcome received by the first member of the team. A subject would undertake the second task with the same partner if both members' minimum claims stated in the questionnaire were compatible with the randomly selected allocation rule.

We invited subjects through campus-wide posters and e-mail advertisements to participate in experimental sessions that lasted approximately ninety minutes. In order to perform an international comparison of the role of social norms in team formation experiments we recruited subjects both at a Spanish University in Navarra and at a Japanese University in Osaka. Spanish and Japanese people a priori differ in cultural aspects. For example, the Japanese corporate culture is associated with the extensive use of teamwork (Abegglen 1958, Haitani 1990, Koike 1988). Japanese organizations have actually served as a model to US firms in order to develop the use of teams in the workplace.

IV. Equality versus equity: a comparison of Japanese and Spanish subjects

We briefly present the results of our experimental sessions in the two countries stressing the relative importance of *equity* and *equality* norms.

IV.1. Equality norm and excessive cooperation

The experimental sessions with Japanese subjects involved 32 students recruited at Osaka University. We observe that Japanese subjects claim almost systematically (82% of the time) an equal share of the team outcome even when they largely outperformed their partner. Interestingly, Japanese subjects were able to assess their relative contribution accurately but were not willing to allocate rewards based on individual contributions. As a result, Japanese subjects, by complying with the equality norm, were ready to accept the formation of highly inefficient teams. In that sense, we observe an excess of cooperation in the case of Japanese subjects. Team formation is excessive since the great majority of subjects would actually obtain higher monetary rewards by working alone on the task. Japanese subjects were typically achieving the maximum level of performance whether they worked alone or in a team. Consequently, the *equality* norm, as emphasized by experimental economists in the context of the ultimatum game, is

robust to settings in which subjects bargain over real team tasks outcomes. The *equality* norm praised by behavioral economists and anthropologists (Henrich et al. 2004) as a fundamental element for effective cooperation among individuals can actually backfire and lead to excessive cooperation.

IV.2. Equity norm and self-serving attributions

The experiments with Spanish subjects were performed at Universidad de Navarra with a total of 43 students. The comparison of Spanish and Japanese subjects leads to striking results. First, Spanish subjects were typically less accurate and objective in assessing their own contribution to the team outcome. In line with results from the social psychology literature on biased self-attribution (Bradley 1978, Larson 1977, Miller and Ross 1975, Zuckerman 1979) subjects tended to overestimate their relative contribution to the performance of the team. Indeed, only 18% of the subjects perceived themselves as contributing less to the team outcome than their partner whereas 36% of the individuals estimated their contribution to be higher than their partner's. In addition, Spanish subjects' claims of the team outcome were driven by *equity* rather than *equality*. Indeed, only 41% of the Spanish subjects required an equal splitting of the team outcome in the second task compared to the 82% of Japanese subjects following the *equality* norm. The claims of Spanish subjects were in line with their perceived contribution to the joint outcome of the first task. That is, subjects consistently asked for a higher share of the team outcome when they perceived to outperform their partner. Given that Spanish subjects suffered from self-serving attributions they tended to claim high shares of the team outcome of the second task and this tended to undermine the formation of teams in opposition to the case of Japanese subjects.

V. References

Abegglen, J. 1958. *The Japanese factory*. Cambridge, MA, MIT press.

Andreoni, J., B. Bernheim. 2007. *Social image and the 50-50 norm: a theoretical and experimental analysis of audience effects*, mimeo.

Bradley, G. 1978. *Self-serving biases in the attribution process: a reexamination of the fact or fiction question*. *Journal of Personal and Social Psychology* 36, 56-71.

Fehr, E., K. Schmidt. 1999. *A theory of fairness, competition, and cooperation*. *Quarterly Journal of Economics*, 114, 817-868.

Fischer, R., P. Smith. 2003. *Reward allocation and culture: a meta analysis*. *Journal of Cross-Cultural Psychology* 34, 251-268

Güth, W., Schmittberger, R., B. Schwarze. 1982. *An experimental analysis of ultimatum bargaining*. *Journal of Economic Behavior and Organization* 3, 367-388.

Haitani, K. 1990. *The paradox of Japan's groupism: threat to future competitiveness?* *Asian Survey* 30, 237-250.

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., H. Gintis. 2004. *Foundations of human sociality: economic experiments and ethnographic evidence from fifteen small-scale societies*. Oxford University Press.

James, K. 1993. *The social context of organizational justice: cultural, intergroup, and structural effects on justice behaviors and perceptions*. In R. Cropanzano (Ed.), *Justice in the workplace: approaching fairness in HRM*, 21-50. Hillsdale, NJ: Lawrence Erlbaum.

Koike, K. 1988. *Understanding industrial relations in modern Japan*. New York, St Martin's.

Larson, J. 1977. *Evidence for a self-serving bias in the attribution of causality*. *Journal of Personality* 45, 430-441.

Miller, D., M. Ross. 1975. *Self-serving biases in the attribution of causality: fact or fiction?* *Psychological Bulletin* 82, 213-225.

Roth, A., Prasnikar, V., Okuno-Fujiware, M., S. Zamir. 1991. *Bargaining and market behavior in Jerusalem, Ljubljana, Pittsburgh and Tokyo: an experimental study*. *American Economic Review* 81, 1068-1095.

Schurter, K., B. Wison. 2007. *Justice and fairness in dictator games*. mimeo.

Zuckerman M. 1979. *Attribution of success and failure revisited, or: the motivational bias is alive and well in attribution theory*. *Journal of Personality* 47, 245-285.

Ryan Muldoon
University of Pennsylvania

Agreeing to Disagree: How Conflicting Norms Can be Mutually Reinforcing

While it has been well-established that social norms can be treated as equilibrium selection devices¹ in social situations, there has not been a corresponding effort to investigate what happens when multiple conflicting norms exist in the same population. Though the literature often posits that social dynamics will or at least ought to lead to a single social norm, this often does not happen in practice. As a case study of a case in which a population does not converge on a single norm, this paper concerns itself with the norms of distributional fairness. Mere reflection on experience shows that there is not a single norm of distributional fairness in the West, but rather, a small number of prominent norms. The three dominant accounts of fairness are to each equally, to each according to merit, and to each according to need. Each of these basic conceptions of fairness compete for being the norm of fairness. Using this case study, we can see how a lack of convergence can be good for society. But first, we must understand the dynamics of the norms of fairness themselves in order to motivate this claim.

While it would be incorrect to say that it was inevitable that the three options described would become the prominent accounts of fairness, we can say something about whether it is desirable that society is left with three norms rather than one. On the face of it, this would pose a problem. To see why, let us suppose that people trade with each other, but if they are to trade, they must decide how to distribute the gains from trade. If they are able to settle on an agreement, they can trade, and distribute their gains in a mutually-agreeable manner. If they cannot agree, then they do not trade.² Intuitively we can see that in populations with multiple conflicting fairness norms, people will at least sometimes disagree about what a fair distribution would be, whereas in a population with a single norm, people would never disagree. Thus, the homogeneous population would be more efficient: every potential transaction would happen. As we increase the number of competing norms in the population, the likelihood of people disagreeing increases, which decreases social efficiency.

In an evolutionary context, less efficient populations are less evolutionarily stable, as mutants that increase efficiency will displace less efficient agents. In this instance, as more homogeneous populations would have increased efficiency, any stochastic variation that allowed one norm to have a larger share of members would be reinforced under a replicator dynamic. If we view norms as game-theoretic strategies³, each norm would guarantee a payoff against itself, since it would always agree with itself, but there would be instances against other strategies that would have a zero payoff due to disagreement. This would reward the strategy with the largest number of adherents, and would eventually drive out the other strategies.⁴

So it would seem from our preliminary analysis that there is a strong evolutionary pressure forcing populations with multiple norms to fixate on a single norm. In fact, so long as we assume that fairness norms have no effect on the amount to be distributed, it does not matter *which* norm it is any norm would be maximally efficient if it were universal. It is for this reason that many norms can be treated

1 See Bicchieri's *The Grammar of Society* (2006) for a developed account of this idea.

2 This kind of interaction is represented as an ultimatum game in the models developed in the paper.

3 For the remainder of the abstract I use norm and strategy interchangeably. On this model, all norms are treated as game-theoretic strategies.

4 This analysis leaves out relevant details such as the precise payoff structure, and assumes that all strategies do equally well against themselves and against each other. These details are fully specified in the paper-length discussion.

as equilibrium selection devices: on that analysis, it does not particularly matter which equilibrium is chosen, so long everyone coordinates on the same one.

Let us assume that we should have a single norm of fairness, for the efficiency claims made above. Though all of the potential norms of fairness are equally efficient in the sense that everyone will always accept proposed distributions of the gains to trade, they do not have the same distributional properties. So how to choose? We can start by considering the simplest norm: equality. Equality has a nice symmetry to it, and it is easy to measure. However, straight equality might not be as efficient as it first seems: if we suppose that different people have different levels of talent, education, or drive, and that those with more of these qualities can do more with their resources than others can, then it would be more efficient to provide them with more resources. If the more meritorious have more resources, they can in turn generate larger surpluses for society to distribute. So then on efficiency grounds, we ought to abandon equality, and turn to a notion of meritocratic equity.

Alternatively, we could start from equality, and suppose that some agents require more resources to achieve the same levels of life satisfaction. These individuals might be disabled in some way, or have environmental disadvantages. Through no particular fault of their own, these individuals need more than an equal share of goods to have an equal share of life satisfaction. So while our goal was an equal division of goods, this appealed largely because it provided equality amongst members of society. If we find ourselves in a situation in which it takes an unequal distribution to arrive at equal life satisfaction, then we might find reason to abandon equality and turn to a notion of need-based equity.

If we choose one of these two equity-based norms of fairness, a new problem arises. Under a norm of equality, it is very easy to perform the distribution — just divide the goods in two. But equity norms require information: we need to be able to assess who is more meritorious, and who is needier. Since there is no central registry of such facts, we must supply this information for ourselves, and assess what we receive from others. As soon as fairness decisions are based on signals passed by other agents, however, there is an opportunity for cheating. If we assume that we are in a merit-based society, everyone who is not talented has a reason to lie about being talented so as to benefit from the extra goods that talented people receive. People might buy fake degrees online, or buy phony class rings, and thus pay a fairly small cost to be able to present a credible signal of their talent, and reap the rewards, such as a higher salary. The challenge that a homogeneous equity-based society faces is that everyone who is not talented will pay a small cost to pretend that they are, so long as we assume that everyone is rational. The distributional goals of the equity norm are thus subverted: the truly talented do not get more than anyone else, and society ends up back at a roughly equal distribution of goods. There is an efficiency loss, however, because the less talented in society would have all had to spend something on cheating, which is a transaction cost that did not exist with a norm of equality. Of course, the same problem of cheating would arise with a need-based merit. People could instead invest in some signal to fake their disability or other disadvantage. As before, the distributional goals of the norm would be subverted.

So though we have reasons for wanting to move away from a norm of equality, we find is that with equity-based norms, homogeneous populations are taken over by parasitic cheating strategies that ultimately undermine the distributional goals of the norms. This is similar to the effect of crop monocultures: it becomes very easy for parasites to destroy the crop, because the cost of adapting to the particularities of the crop itself are very low compared to the expected gains. Interestingly enough, this comparison to crops provides us with the way to solve the social problem we face. Parasites become a much smaller risk when there is more than one kind of crop growing. If we drastically increase the number of kinds of crops, the risk of parasites goes towards zero. Similarly, if we were to increase the

number of norms in a population, the expected value of cheating on any particular signal drops to zero, since the likelihood of interacting with someone following the appropriate norm goes to zero. So, the small positive cost of cheating overwhelms the expected benefits. As such, no one would choose to cheat.

Populations that have a massive number of conflicting norms would be very successful at driving out cheaters, but it is at the cost of efficiency. Completely driving out cheaters with a large number of conflicting norms would also result in a large number of failed interactions. Thus, it appears we have two different evolutionary pressures: one pushes the population to have fewer conflicting social norms so as to increase efficiency, and the other pushes the population to have more conflicting social norms so as to drive out cheating. It seems that the optimal number of social norms in a population is neither one nor many, but few.

The first half of the paper's analysis provides a detailed evolutionary game theoretic model of the dueling pressures of efficiency and cheater prevention. As a part of this analysis, I offer a robustness check on the results, particularly the effects of different network structures. This descriptive portion of the paper reveals the conditions under which self-interested rational agents can achieve optimal social outcomes, both in the sense that social efficiency is high, and in the sense that the distributive goals of the social norms are met. It is the latter notion of social optimality that we need to pay particular attention to, as this is where the analysis is particularly novel. Namely, if we assume that people at least in part adhere to a particular norm because they believe that the distributive goals of the norm are important, then this is a measure we need to pay attention to.

What is particularly striking is that we find that since the presence of other norms reduces the instances of cheaters being parasitic on a particular norm, the presence of those other norms actually enhances the distributive success of that norm. That is, having people around that disagree with you can enhance your ability to accomplish what you want. The mere act of disagreeing mutually reinforces the interests of both sides.

The second half of the paper seeks to explore the normative consequences of this unusual fact. In particular, we examine the potential institutional frameworks that could be employed to buoy this phenomenon. We begin with an analysis of what structural features of a social network might enhance the stability of a mixed-norm population, and find that networks with sparsely connected clusters are most able to provide stable populations. Being that these network structures are most conducive to maintaining several conflicting norms, to the benefit of the goals of each norm, societies have rational reasons for adopting such network structures. Going beyond the technical analysis, we draw wider conclusions about the consequences of this project. Namely, not only do we have a liberal obligation to tolerate others, but we often have a rational obligation to encourage the flourishing of our ideological opponents, so we might be able to flourish ourselves.

The emergence of cooperation in a heterogeneous world.

An evolutionary approach*

Alessandra Smerilli[†], University of East Anglia

Civil life is essentially a matter of cooperation. Neoclassical economic theory propounds a highly parsimonious view of cooperation as deriving exclusively from the calculation of individual interest. It follows from this parsimonious view that a person would never cooperate in a non-iterated prisoner's dilemma game. Should one then observe in the laboratory that some players cooperate even in a one-shot game, the explanation is simple: they played the 'wrong game' or, simply, they were irrational. This is the thesis put forward, for example, by Binmore: "the framing of the game triggers a social norm that players are accustomed to using when going about their everyday affairs" (Binmore 2006, p. 85). It is for this reason that agents may cooperate in a non-iterated prisoner's dilemma. If instead the game is repeated, the traditional theory justifies the cooperation simply by citing self-interest (this being the so-called 'folk theorem') or contracts with enforcement.

In reaction to this excessively parsimonious view of cooperation, recent years have seen development of a body of literature ('social preferences' theory) which instead seeks to explain why even in a one-shot non-cooperative game (i.e. the 'ultimatum' or 'trust game') it may be rational to play 'cooperatively'. The explanation, of which there are several variants, rests on the idea of a psychological payoff: in some types of interaction, certain non-tangible factors (for instance, an inequality aversion, reciprocity, etc.) may change the game's payoff structure. It is as if, besides the agent's utility function, there are further components (of a non-monetary kind) explaining the emergence of cooperative behaviour in contexts where the standard neoclassical theory would exclude it.[‡]

This is the explanation of cooperation advanced by behavioural economists (see Gintis (2004) and Bowles and Gintis (2004)), who base their analyses of cooperation on the theory of strong reciprocity (Fehr and Gächter (2000)). By 'strong reciprocity' they mean a social norm which, in a manner costly to the individual, rewards those who behave well and punishes those who behave badly. This theory of cooperation stands in methodological and cultural opposition to the mainstream economic theory: whereas standard economics (i.e. that of Binmore, 2005) envisages

* A previous version of this paper has been written with Luigino Bruni.

[†] PhD Student in Economics at University of East Anglia.

[‡] For a survey see Smerilli 2006.

nothing but self-interest and monetary incentives, strong reciprocity theory explains the emergence of cooperation on the basis of a form of altruism which does not even require the game's repetition.

In this paper I adopt a different perspective. I propose a theory of cooperation which is anthropologically more generous than that of standard economics, but I do not embrace the strong reciprocity thesis. I put forward a pluralistic and multidimensional view of cooperation and consequently examine aspects hitherto insufficiently explored by economic and social theory. Specifically, I seek to show that, in certain settings, less 'altruistic' forms of cooperation may combine with more gratuitous ones.

I accordingly construct dynamic models which will enable us to analyse diverse patterns of cooperation or reciprocity. There are many such patterns, not all of them based on self-interest, but all of them important for understanding the dynamics of civil life.

I base the analysis on the Prisoner's Dilemma (PD) game, which is widely used to analyse cooperation because it lends itself well to the modelling of 'difficult' cooperation: the kind that occurs in situations where there is no enforcement and where there is always an incentive for non-cooperation. I believe that these situations are frequent and relevant – although in civil society individuals play many games, not only the PD – and that they are important in the real dynamics of cooperation in civil life.

In section 2 I analyse the evolution of cooperation in a 'one-shot' context, while in section 3 I apply the evolutionary analysis to repeated games. In section 4 I concentrate on analysis of situations in which four strategies interact (Gratuitous, Braves, Cautious and Nasty), also furnishing simulations. The paper concludes with a brief discussion on the results of our analysis. Main conclusions are the followings:

- (a) *The 'crucial' role of Gratuitous (G) types.* in my analysis G types should not be too numerous, because if they are they compromise themselves and also the survival, for example, of Braves (B). In populations where non-cooperation is possible (which is the case of all real ones), unconditional acts are essential, but when too numerous, they become counter-productive.
- (b) G types perform a vital role, for only they can activate the cooperation of Cautious (C). Without the presence of G types, Cs would never experience cooperation and therefore would never respond with an act of cooperation. G types are consequently valuable, but they should be protected. The success of numerous forms of cooperation – from firms to families – depends also, and sometimes above all, on the presence of a small number of unconditional reciprocators able to activate people who would never be so activated if they only interacted with conditional cooperators.

(c) *Alliances: C types*. These are ‘activated’ by Gs, but at the same time their presence is highly beneficial to Gs because it increases their expected utility. Gs, in fact, cooperate with Bs and with Cs, but they are exploited by Ns. In a four-strategy world, Cs protect the Gs against extinction.

Cooperation is therefore favoured by heterogeneity. From a mathematical point of view, it might be objected that G types are not necessary. The onset of cooperation would only require slightly more sophisticated Bs. But this was not the purpose (i.e. to study which strategies favour cooperation) for which the model was conceived. The analysis started from the assumption that behaviours like G exist in civil society. (And who could deny the presence in the real world of unconditional actions? Even Binmore (2006) with his orthodoxy and anthropological parsimony admits their existence). The model I present, has sought to analysis the conditions under which unconditional actions can not only survive but also perform a virtuous civil role.

REFERENCES

- Antoci A., Sacco P. e Zarri L. (2004) “Coexistence of Strategies and Culturally-Specific Common Knowledge: An Evolutionary Analysis”, *Journal of Bioeconomics*, vol. 6, pp. 165-194
- Axelrod R. (1984) *The evolution of cooperation*, Basic Books Inc., Publishers, New York
- Binmore K. (2006) *Natural Justice*, Oxford University Press, USA
- Binmore K. (2006) “Why do people cooperate?”, *Politics, Philosophy and Economics*, vol. 5 (1), pp. 81-96.
- Bomze I. (1983) “Lotka-Volterra Equation and Replicator Dynamics: A Two-Dimensional Classification”, *Biological Cybernetics*, vol. 48, pp. 201-211.
- Bowles S. e Gintis H. (2004) "The evolution of strong reciprocity: cooperation in a heterogeneous population." *Theoretical Population Biology*, vol. 65, pp. 17-28.
- Bruni L. (2006) *Reciprocità. Dinamiche di cooperazione, economia e società civile*, Bruno Mondadori, Milan.
- Bruni L. and Crivelli L. (2004), *Per una economia di comunione - un approccio multidisciplinare*, edit by, Città Nuova, Rome.
- Bruni L. and Smerilli A. (2004), “I dilemmi dell’individualismo e il paradosso della reciprocità. Ipotesi e giochi”, in Bruni and Crivelli (2004).
- Fehr E. and Gächter S. (2000) “Fairness and Retaliation: The Economics of Reciprocity”, *Journal of Economic Perspectives*, 14, pp. 159-181.
- Gintis H (2004) “Modeling Cooperation Among Self-Interested Agents: A Critique”, *The Journal of Socio-Economics*, 33, pp. 311-322.
- Heckathorn D. (1996) “The dynamics and dilemmas of collective action”, *American Sociological Review*, vol. 61, pp. 250-277.
- Hirshleifer J., Martinez Coll J. (1991) “The limits of reciprocity”, *Rationality and Society*, vol. 3, pp. 35-64.
- Smerilli A. (2006) *Comportamenti cooperativi e razionalità del noi (we-thinking)*, PhD thesis, Università La Sapienza, Roma.
- Sugden R. (2003) “The logic of team reasoning”, *Philosophical explorations*, vol. 6, pp. 165-181.
- Sugden R. (2004) *The economics of rights, cooperation and welfare*, second edition, Palgrave Macmillian, London.

The logic of collective sanction

Benoît Dubreuil

Research fellow, FNRS, Université libre de Bruxelles

Abstract

One typical problem in social science concerns the interpretation of individual agents in persistent situations of injustice. Social scientists have historically hesitated between two types of explanation. The first can be qualified of “culturalist” (or “ideological”) and stress the importance of shared social norms in ensuring social cohesion. From this point of view, agents do not oppose to injustice because their normative expectations differ from ours. The second type of explanation is associated with rational choice theory and interprets the persistence of situations of injustice by referring to collective action problems. People do not oppose injustice because collective action is too costly and agents are unable to control free-riders.

These two types of explanation face important difficulties. On the one hand, culturalist and ideological explanations often propose uncharitable interpretations of the behavior of oppressed individuals, by ascribing them irrational normative expectations. On the other hand, rational choice explanations have traditionally failed to explain how agents can in certain situations (e.g. revolts, revolution) overcome persistent problems of collective action, achieve mass mobilization, and overthrow oppressive regimes.

I propose that the problem can be solved by paying more attention to the interaction between expectations, emotions, and motivations. From the culturalist viewpoint, social change and revolt happen when normative expectations are violated. The problem, as rational choice theorists argue, is that individuals in situations of injustice often do not share the same normative expectations and that the oppressed already consider the situation as “unfair” before to revolt. But if the normative expectations are already violated, what can explain sudden mass mobilization and revolt?

On the basis of recent research in experimental economics, I argue that the urge to punish is not triggered by the violation of normative expectations. Indeed, we already expect people to be unfair to a certain extent, and do not get angry when they do. What really make us angry is when people are unexpectedly unfair. This triggers the powerful emotion of anger, transforms our utility function and explains mass mobilization to realize collective sanction. In sum, explaining the persistence of injustice and sudden social change is not possible on the basis of normative expectations and the structure costs and benefits associated with collective action. Social scientists should also pay attention to the way non normative expectations frame the triggering of emotions and transform agents’ utility function.

Experimental background

In a series of experiments, van Winden, Reuben, and Bosman proposed to clarify what emotion underlies punishment (reviewed by van Winden 2007). The experiment is based on the “power-to-take game.” In this game, two players are given an initial sum of money (5 MUs). One subject, the proposer, has the power to take a part of the money of the other. In a second turn, the responder has the possibility of destroying a portion of his income, thereby incurring costs for himself but simultaneously reducing the benefits of the proposer.

In contrast with many experiments in behavioral economics, van Winden and colleagues did not simply speculate about fairness but asked participants to specify what take rate they consider to be fair. At the same time, they evaluated the expectations of participants concerning proposers’ take rate. Those are obviously two different questions: participants typically fix the “fair take rate” much lower than the “expected take rate”. More prosaically, they do not expect the proposers to be entirely fair. The interesting result of the power-to-take experiment is that the expected take rate is a much better predictor for anger and punishment than the fair take rate. Responders get angry when proposers do not conform to expectations; that is, they get angry in response to “takes” that are higher than expected rather than higher than what they consider to be fair (van Winden 2007: 43). The role of anger is well-supported by the fact that the decision to punish is strongly time-sensitive. Responders who choose to destroy nothing or everything decide pretty fast, while those who choose to destroy only a part of their own income take more time. This can be easily explained by the pattern of emotional arousal associated with anger (van Winden 2007: 42).

Another noteworthy result is the asymmetric nature of decision-making among proposers and responders. In repeated power-to-take games, judgments of fairness do play a role for proposers, by eliciting the emotions of shame and guilt. Proposers who reported a high level of shame and guilt in the first round of the game were more likely to lower their take rate in the subsequent round. Moreover, this effect was particularly strong among proposers whose chosen take rate exceeded what they considered to be the fair take rate. It was also stronger among those who had been punished in the first round (Reuben and van Winden forthcoming; van Winden 2007). The conclusion that we can draw is that the emotions of shame and guilt are elicited by the joint effect of effective sanctions and judgments of fairness.

The experiments described above explore sanctions among players who have direct stakes in the outcome. In real life, by contrast, a lot of sanctioning comes from third parties or individuals who have nothing to gain or lose in the interaction. Few experiments have dealt with third party punishment, but their results are telling. In one set of experiments, Fehr and Fischbacher (2004) have measured third party punishment in dictator and prisoner's dilemma games. They discovered that third party punishment was real, although it was less important than second party punishment. Thus, the fact of being directly harmed by an action arguably elicits more anger than simply seeing someone harmed. Fehr and Fischbacher conclude that punishment by a single third party is insufficient to make norm violation unprofitable, a problem the presence of several third

parties can remedy. This feature of sanction, I will argue in chapter 4, creates a pressure for increased group size in humans.

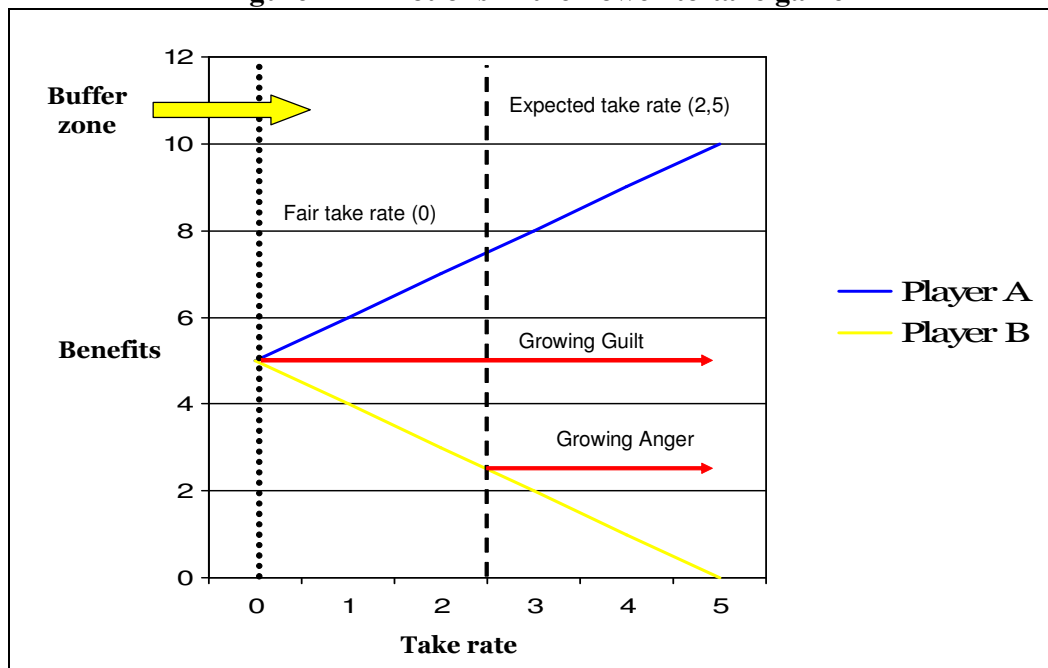
Carpenter and Matthews (2005) proposed a second set of experiments to explore third-party punishment. They used public good games with multiple groups and the possibility of sanction within and across groups. They wanted to address one weakness of Fehr and Fischbacher's experiments, where all that the third party is allowed to do is punish. If the third party's participation is restricted to norm enforcement, there are reasons to expect him to spend more on punishment than otherwise. Carpenter and Matthews (2005) confirm that third parties do punish, although they tend to pay much less for punishment than second parties.

Carpenter and Matthews (2005) suggest that third party and second party punishments have different emotional bases. They borrow Elster's description of the circumstances that elicit both emotions: "if I believe that another has violated my interest, I may feel anger; if I believe that in doing so he has also violated a norm, I feel indignation." (Elster 1998: 48). If this is right, the decision to punish would be based primarily on anger in second parties and on indignation in third parties. This distinction makes good sense, although it is not clear that anger is triggered by the violation of one's interest, as proposed by Elster. As I argued above, the fact that someone is more unfair than what we expected is the best predictor of anger in social dilemmas, and not the violation of interests (van Winden 2007).

To summarize, we can draw an interesting parallel between the emotions of guilt and indignation. These two emotions are triggered by the violation of a norm of fairness and elicit a motivation to provoke a change in behavior. Guilt draws one's attention to one's own actions and motivates her to repair past violations. Indignation attracts one's attention to others' actions and motivates sanction. Anger, on the other hand, is associated with the urge to punish someone whose unfair behavior goes beyond what we expected. The violation of normative expectations can trigger different emotions depending on one's position in a social interaction. Guilt is dominant in appropriately punishing first parties, anger in punishing second parties, and indignation in punishing third parties.

One conclusion that can be drawn from this picture is that the emotions underlying effective punishment are not triggered by the violation of the same expectations. Behavioral expectations are usually set below expectations in terms of fairness and, consequently, anger will tend to be harder to elicit than guilt and indignation. Figure [x] represents the gap between the triggering points of guilt, indignation, and anger in a classical ultimatum game. In this game, fair transfer will usually be set at 50%, but expected transfer at about 25%. I will describe transfers in the range between 25-50% as a "buffer zone" where guilt and indignation are elicited, but not anger. In the buffer zone, the probability of sanction is low, because indignation is much less effective than anger in motivating punishment. At the same time, punishment, if it occurs, can already be effective, because the emotion of guilt is already activated.

Figure 1 – Emotions in the Power-to-take game



The existence of such a buffer zone should not be surprising. As I argued above, punishment can only foster cooperation if it is perceived as fair. At the same time, individuals tend to adopt self-serving views about fairness. Consequently, if anger was triggered by the violation of expectations about fairness, there would be a much higher probability of punishment in the absence of guilt and, therefore of retaliation and escalation of violence. Our everyday interactions would be much more violent and punishment would be no more beneficial to cooperation. In real-life, however, experiencing guilt is often easier than experiencing anger, although there is probably much heterogeneity across people. We wonder what we could have done wrong even when others show no sign of anger at all. I suggest that we understand this feature of our psychology as essential for the stability of social interactions, without which punishment would not be associated with discipline but with the rule of the stronger.

References

- Carpenter, J. P. and P. H. Matthews (2005). Norm Enforcement: Anger, Indignation or Reciprocity? *IZA Discussion Papers*, 1583.
- Elster, J. (1998). Emotions and Economic Theory. *Journal of Economic Literature*, 36, 47-74.
- Fehr, E. and U. Fischbacher (2004). Third Party Punishment and Social Norms. *Evolution and Human Behavior*, 25, 63-87.
- Reuben, E. and F. van Winden (forthcoming). Social Ties and Coordination on Negative Reciprocity: The Role of Affect. *Journal of Public Economics*.
- van Winden, F. (2007). Affect and Fairness in Economics. *Social Justice Research*, 20(1), 35-52.

Reciprocity, Exchange and Redistribution.
An experimental investigation inspired by Karl
Polanyi's
The Economy as Instituted Process

GIUSEPPE DANESE*

LUIGI MITTONE

CEEL - Computable and Experimental Economics Laboratory

Department of Economics

University of Trento

Via Inama, 5 38100 Trento Italy

EXTENDED ABSTRACT (1998 words)

Key Words: Reciprocity, Redistribution, Exchange, Comparative Institutional Analysis.

JEL codes: Z13, D02, C91.

* Corresponding author. Telephone: +39.461.882246; fax: +39.461.882222.

E-mail address: gdanese@economia.unitn.it (G. Danese).

INTRODUCTION

The theoretical background of this essay comes from Polanyi's investigation on three forms of allocation: reciprocity, redistribution and exchange. In his own words, "reciprocity denotes movements between correlative points of symmetrical groupings; redistribution designates appropriational movements toward a center and out of it again; exchange refers here to vice-versa movements taking place as between "hands" under a market system. Reciprocity, then, assumes for a background symmetrically arranged groupings; redistribution is dependent upon the presence of some measure of centricity in the group; exchange in order to produced integration requires a system of price-making markets" (Polanyi, 1957, 250).

The meaning of the three allocation modes is better understood as expression of underlying social relationships: kinship and status for reciprocity, where the key reference is to Malinowski's *Argonauts* (2004). In the interpretation given by Polanyi, the *kula* there described is centred on the act of giving as valuable in itself, without the need of any means-ends fit type of reasoning: "Trobriand economy [...] is organised as a continuous give-and-take, yet there is no possibility of setting up a balance, or of employing the concept of a fund. Reciprocity demands adequacy of response, not mathematical equality" (Polanyi, 1957, 273). The market is a unique transactional mode, given that it is *disembedded* from the social matrix.

Polanyi states clearly that the forms of integration he describes cannot be considered as projections of personal attitudes at an aggregate level. With reference to reciprocity and redistribution, the presence of well-identified social norms, respectively symmetry and centricity, are necessary in order to produce integration. The prediction he makes is that only in

symmetrically organised groupings will reciprocative behaviour result in economic institutions of some historical and anthropological importance (Polanyi, 1957, 251).

Our main objectives in this experimental investigation is reproducing in a laboratory setting two of the three Polanyian circuits, holding the experimental settings as homogeneous as possible throughout. We do this in order to compare the allocative-efficiency levels of the circuits. Allocative efficiency is here measured as the ratio of the sum of the actual gains of the players in the game, to the potentially attainable gains if optimal consignment-decisions take place (Gode & Sunder, 1997).

With reference to our theoretical startingpoint, we tried to isolate out of that apparatus an element of comparative institutional analysis among different interaction structures. Our hypothesis is that while the market can function well even in the absence of forms of induced socialisation, for gift-trade rings to prompt cooperation there need to be in place institutional refinements promoting forms of symmetry-acknowledgement.

RECIPROCITY

The subjects played an investment game with indirect reciprocity, in which the pile of points available, in the form of endowment, rises at a steady rate, in the case in which trust and reciprocity prevail consistently (centipede game-like). Our game shares some features with Greiner & Levati's (2003). We formed randomly groups of ten players, students of the University of Trento. The game works as follows: the first player receives from the experimenter a very small amount of points (16 eurocents in Treatment 1, and 28 in Treatment 2). The order of play was random and all the choices anonymous. The first player could decide how much to keep for himself of the endowment and how much to send to a generic "next player".

The amount sent was multiplied by a factor of 2 by the experimenter. Player number 2 took a similar decision, with the multiplication of points taking place at each gift-decision of the players. In our experiment, player 10 is only a dummy, for the amount that reaches him is divided among all group members in equal shares. The final payoff of the players was calculated as the simple sum of the points they decided to withhold, and of the points that reached player number 10, divided by 10. The players' consignment-decisions were constrained: they could not send zero. Still, they could choose to adopt a smallest-granularity consignment-decision, i.e. sending 1 point only. Players number 2 to 9 had a further option: they could decide to end the experiment. In the case of closure, the player's endowment is divided in equal shares between himself and all the previous players, with the remaining players obtaining zero. Players saw on the screen, when they had to make their decision, the amount of points that *on average* the previous players had available. All players were informed in the instructions of the amount of points they could have gained in the case in which all players sent all the points they had available.

We have conducted a baseline and a pre-play communication treatment, the difference being whether the players were allowed a stage of pre-play communication.

The main results of the baseline experiment are shown in Figure 1.

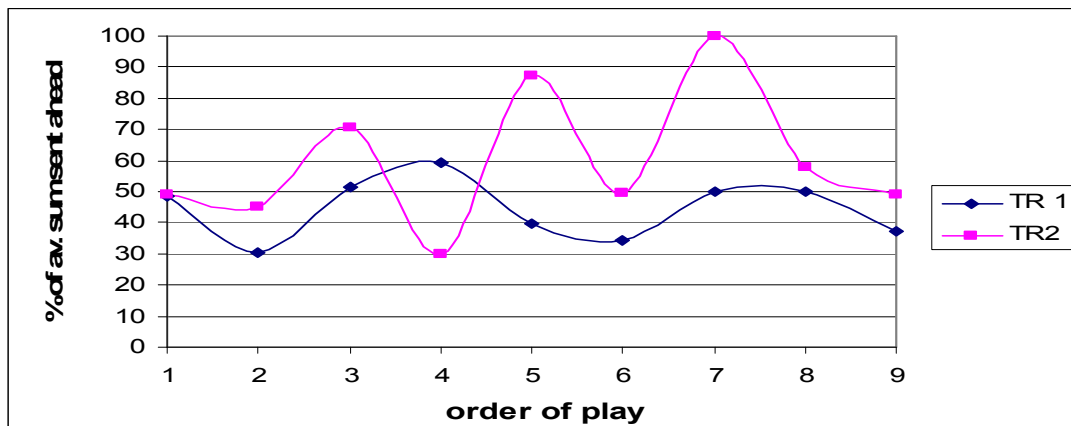


Figure 1. Baseline experiment: gift-decisions (as percentage of available sum, average figures)

Figure 1 shows that players rarely send the whole sum available, and that decisions to send ahead a significant share of the endowment are usually followed by decisions to withhold a relevant part of the available sum. Overall, the baseline experiment has yielded modest levels of allocative efficiency: in treatment 1, 0.48 %, in treatment 2, 1.7% (average figures). The difference across the two treatments is the amount of points available to player number 1 ("LOW" in Tr.1, "HIGH" in Tr. 2). Looking at the strategies used by the players, there is a prevalence of "indirect tit for tat" (31.7%), whereby the player reciprocates forward exactly the same amount of points the previous player sent her, and "unkindness" (38 %), whereby the player reciprocates less than what she has received.

The players report in the debriefing survey that the fair share of the endowment to send to next player is 50%. Interestingly, the players are usually consistent, i.e. they have taken a consignment-decision during the game which mirrors well the perception of fairness they express in the debriefing (58 % of the cases with an interval of $\pm 10\%$; 72% of the cases with an interval of $\pm 20\%$). Furthermore, there seems to be generalised lack of trust towards the last players' consignment-decision (68%).

In the pre-play communication game we asked the experimental subjects to discuss for 5 minutes about three questions we provided them with. The experimenters were not present during the discussion, nor did they record it, as they realised that doing so would have produced undesirable effects on the verbal behaviour of the players. The questions were the following:

1. according to you, if all the players send half of the sum they have available, how much will all participants win?

Through this question, we wanted to prompt reasoning by the player as to the benefits arising out of sending all the amount of points they had available.

2) do you think that the introduction of a rule of behaviour among yourselves could be helpful in order to raise the payoffs of all participants?

This question was meant to elicit the possibility of an shared pact among the players, thanks to which a Pareto-efficient outcome could be achieved.

3) do you think that this rule of behaviour that you have just discussed will be used by the players during the game? I would like to remind you that your choices will be anonymous and free.

This question was meant to prompt reasoning about the unequal positioning of the players, given that if all players sent the whole amount available, player number 9 would have had a relevant sum at her disposal. Given that players were discussing behind a veil ignorance, this question was meant to raise the consciousness that all players faced a common weakness, arising out of the compliance problem with the agreed-upon rule of behaviour of those playing last.

A glimpse at the gift-decisions of the players shows that the cheap-talk stage has favoured the players' willingness to send (Figure 2).

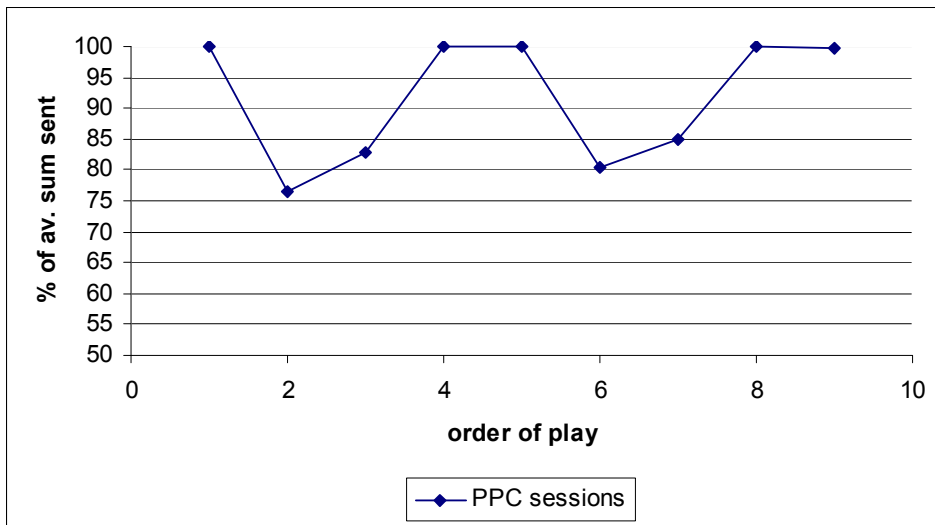


Figure 2. PPC variant: gift-decisions (as percentage of available sum, average figures)

The average allocative efficiency of the PPC sessions is 28.20%, versus 0.48% of the baseline experiment with tr. 1, which constitutes our reference point, given that all PPC sessions were conducted under treatment 1 (i.e., with a “low” initial endowment). Furthermore, the kindness strategy overall prevails (87%). The debriefing survey confirms that the cheap-talk stage has modified the system of beliefs of the players. They report that the fair share to send ahead is 100%. Usually players are consistent with their choices (58 % with $\pm 10\%$ interval, 66% with $\pm 20\%$).

We have then tried to explain the effectiveness of cheap-talk in promoting players’ willingness to send. Our hypothesis is that cheap-talk has eased the acknowledgement of a form of symmetry shared by the players. In particular, we designed cheap-talk topics-for-discussion in such a way as to raise players’ consciousness that all of them, before playing the game, behind a veil of ignorance, shared a common condition of weakness, arising out of everyone’s potential opportunism in the final stage of the game. The debriefing-survey confirms the usefulness of the communication stage in

order to increase the willingness of the players to send points (84%). The subjects who state in the debriefing that they would have behaved opportunistically, were they player number 9, go as down as 55% of the sample (vs. 68% of the baseline experiment).

EXCHANGE

The market exchange game was designed as a bargaining process, with an equal number of "buyers" and "sellers" (5+5). Sellers decide how much to place on the market and a division of the sum, knowing that the experimenter doubles the amount on the market. Buyers accept (or not) one of the offers within a maximum of 4 rounds. Both sellers and buyers are informed that they need to conclude one transaction in order to earn a positive payoff. We have tested two treatments, one in which buyers are informed of the sum available to buyers, and one in which this piece of information is not given.

Restricting the interaction to two players has increased efficiency dramatically (75%), regardless of the information package available to "buyers" and "sellers".

FINAL REMARKS

Concluding, economic anthropology provides valuable hermeneutic tools in order to interpret the role of cheap-talk, as an experimental coordination-aid. Furthermore, systems of indirect reciprocity among anonymous players can seldom function in the absence of definite institutional refinements, promoting forms of symmetry-acknowledgement. This result stands in marked confirmation of the anthropological literature,

coming from Polanyi's idea of symmetry, as well as Marcel Mauss' idea of gift-exchange as a *fait social total*.

Among the possible improvements of the experimental design, one of the most relevant would be to try to mimic the "bi-directional" nature of the *kula* trade system. This would imply allowing the players to play more than once or introducing some form of bi-directionality of trade (e.g., the possibility to send experimental points not only from left to right but also from right to left). In this manner, the institutional characteristics of the *kula* trade would be better reproduced in the artificial setting of the experiment and, given our theoretical starting point, this would produce higher levels of allocative efficiency.

REFERENCES

- Gode, D.K. & Sunder, S. (1997). What makes markets allocationally efficient?, *The Quarterly Journal of Economics*, 112, 603-630.
- Greiner, B. & Levati, M.V. (2003). Indirect Reciprocity in Cyclical Networks, an experimental study, *Max Planck Institute for Research into Economic Systems*. Retrieved June 1, 2007, from https://people.econ.mpg.de/~levati/JoEP03NZ072_GreinerLevati.pdf.
- Malinowski, B. (2004) (or. ed. 1922). *Argonauti del Pacifico occidentale: riti magici e vita quotidiana nella società primitiva*. Torino: Bollati Boringhieri.
- Polanyi, K. (1957). The Economy as Instituted Process. In K. Polanyi, C.M. Arensberg & H.W. Pearson (Eds.), *Trade and Market in the Early Empires*. New York: The Free Press.