An ESRC Research Group

Strategy Choice and Cognitive Ability in Field Experiments

GPRG-WPS-034

Abigail Barr, Magnus Lindelow,

Jose Garcia-Montalvo and Pieter Serneels

Global Poverty Research Group

Website: http://www.gprg.org/

The support of the Economic and Social Research Council (ESRC) is gratefully acknowledged. The work was part of the programme of the ESRC Global Poverty Research Group.

Strategy Choice and Cognitive Ability in Field Experiments

Abigail Barr [a 1]
Magnus Lindelow [b]
Jose Garcia-Montalvo [c]
Pieter Serneels [a]

[a] *Centre for Study of African Economies, University of Oxford*
[b] *Development Economics Research Group, The World Bank*
[c] *University of Pompeu Fabra, Barcelona*

December, 2005

Acknowledgements: The experiments and survey upon which this study is based was funded by the World Bank. Abigail Barr's ongoing involvement in the project is funded by the Economic and Social Research Council through the Global Poverty Research Group. We would like to thank Ritva Reinikka and Agnes Soucat for their support and Danila Serra and Daniel Ali for their invaluable work in the field. The findings interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

[1] Corresponding author. Email: abigail.barr@economics.ox.ac.uk Address: CSAE, Department of Economics, University of Oxford, Manor Road, Oxford, OX1 3UQ, United Kingdom.

Strategy Choice and Cognitive Ability in Field Experiments

Abstract: We explore the value of the strategy method to field experimentalists. Specifically, we demonstrate that, while the method may lead to reductions in subject understanding, it also generates valuable insights. We played the Third Party Punishment Game and the Generalized Trust Game with Ethiopian medical and nursing students applying the strategy method to the responding role in each case. Then, making use of two proxy measures for the students' cognitive abilities, we investigate the relationship between strategy-type choices and subject understanding. Thus, we find support for the assertion that apparently random and internally inconsistent strategies are symptomatic of problems of cognition. We also find support for the often, implicitly made assumption that, in BDM-type trust games, the ratio of what is returned to what is sent is an appropriate focus for comparative analyses of responder behaviour. Finally, we find evidence that an observed difference in third party punishing behavior between Swiss and Ethiopian students is due, not to misunderstandings, but to variations in what is perceive as punishable. Our results lead us to conclude that the strategy method is of considerable value in Third Party Punishment Games, but need not be routinely applied in BDM-type trust games.

JEL classification: C93 - Field Experiments; C91 - Laboratory, Individual Behavior

Strategy Choice and Cognitive Ability in Field Experiments

1. Introduction

The strategy method, originally proposed by Selton (1967), is now a well established and frequently used item in the laboratory-based, experimental economist's tool kit. The method involves asking a subject what they would do in each and every possible state of the world, where the latter results from the actions of others or a random mechanism at play within the experiment. Once one state of the world has occurred, the action prescribed by the strategy and relating to that state is carried out and the subject is paid accordingly. The strategy method has several advantages over the usual, sequential decision protocol. It provides more data-points per dollar spent on experimental stakes and generates considerably more information about individual subject's motivations, preferences, and behavioural tendencies. Indeed, some research endeavours (e.g., Fischbacher, Gachter, and Fehr; 2001) have been critically dependent on its use for this reason. It also yields information about what subjects would do in rarely occurring states of the world. Thus, Bahry and Wilson (2003) and Henning-Schmidt, Zhuyu. and Chialiang (2001) have identified a tendency on the part of subjects to reject not only low but also high offers in Ultimatum games. Further, when applied to subjects who are responding to the actions of others, the method greatly simplifies the problem of controlling for the actions of those others when conducting cross-subject comparative analyses.

The method also has its disadvantages. Incentives may be weakened, as each state of the world occurs with a less than unitary probability, and may be variably weakened if subjects assign different probabilities to different states of the world. The method may impact directly on behaviour and, thereby, affect an experiment's external validity. However, the evidence is mixed: using different experimental designs, Brandts and Charness (2000) and Oxoby and McLeish (2004) find no evidence of an impact, while Güth, Huck, and Mueller (2001), Blount and Bazerman (1996), and Brosig, Weimann, and Yang (2003) find that there is an impact. And finally, the method can be difficult to explain to subjects and, so, lead to reduced subject understanding and data quality.

Due to these disadvantages, the strategy method has rarely been applied in field experiments and combined experimental and survey studies. This is lamentable as, especially in combined studies, greater comparability of behavioural measures across subjects would be a considerable advantage. Reduced subject understanding is a particular cause for concern. However, while the method may reduce understanding, it also generates insights into individual subject's understanding that could be of use not only within the context of the study in question but also to researchers planning new studies. To our knowledge, only Bahray and Wilson (2003) have employed the strategy method during a field experiment and, subsequently, drawn conclusions about the relationship between individual subject cognition the type

of strategy chosen. Thus, they argue that U-shaped rejection strategies in Ultimatum games played in Russia are not symptomatic of misunderstandings. However, the strength of their argument depends on the assertion that apparently random and internally inconsistent strategies are associated with problems of cognition and the finding that a different type of subject chooses these as opposed to both monotonically declining and U-shaped strategies.

The aim of this paper is to explore and demonstrate the value of the strategy method to field experimentalists. We build on and complement Bahry and Wilson's (2003) work, by investigating the relationship between individual subject cognition and strategy choice in two games played by medical and nursing students in Ethiopia. The two games were Fehr and Fischbacher's (2004) third party punishment game (TPPG) and Buchan, Crosson, and Dawes' (2002) generalized trust game (GTG), In the TPPG the Ethiopian subjects could be classified into three groups according to the type of strategy they chose: the first chose strategies that are qualitatively comparable to strategies chosen by Swiss undergraduates in the original laboratory experiment; the second chose apparently random and internally inconsistent strategies; and the third chose strategies fitting neither of these characterizations; i.e., strategies that appear to be non-random and internally consistent but inconsistent with the model used to explain the behaviour of the first group. Using two non-experimentally derived proxy measures for individual subject cognition, we investigate the hypothesis that subjects sort across strategy-types with reference to their cognitive abilities and consequent, likely level of understanding.[2] In so doing, we find support for Bahry and Wilson's assertion: in comparison with the subjects in the first group, the subjects in the second appear less cognitively able, suggesting that apparently random and internally inconsistent strategies may, indeed, be symptomatic of misunderstanding. However, we also find that the third group of subjects are statistically indistinguishable in terms of their cognitive abilities from the first, while being significantly different from the second.

We conduct a similar analysis for the strategies chosen in the GTG. However, in this case, while once again testing the hypothesis that subjects sort across strategy-types with reference to their cognitive abilities, we are particularly interested in an assumption often, but never explicitly, made by researchers when comparing reciprocating behavior across subjects: that each subject's underlying tendency to reciprocate is captured by the ratio of how much they return to how much they receive and that this ratio does not vary depending on how much they receive. This assumption is valuable, especially to field experimentalists interested in explaining behavioural diversity, because it facilitates cross subject comparisons even when the strategy method is not applied and the subjects in question receive different

[2] Our endeavour differs markedly from that of McKelvey and Palfrey (1995, 1998) and Bolton and Ockefels (2000). They estimate single decision error parameters and, in the case of the latter, a single inequality aversion parameter for entire samples of subjects using only experimental data. In contrast, we focus on subject diversity with respect to decision error and make use of both non-experimental and experimental data to study its impact on strategy choice.

amounts. If this assumption is correct, under the strategy method, subjects should choose strategies that, when plotted in received-returned space, are straight lines passing through the origin. A significant proportion of our subjects do not choose such strategies. However, we find evidence that deviations from such a strategy-type are associated with lower cognitive ability and, hence, a greater likelihood of misunderstanding.

The paper has seven sections. Following this introduction, section 2 briefly describes the context of the study. Then, in section 3 we outline the two experimental games and the design of our experiment. In section 4 we describe our Ethiopian subjects and introduce our two proxies for subjects' cognitive abilities. In sections 5 and 6 we present and analyse the strategy data derived from the TPPG and the GTG respectively. In section 7 we conclude with a discussion about the implications of our results, especially for field experimentalists.

2. The context of the study

Every year thousands of young professionals join what the World Bank (2003) refers to as the 'frontline' in development and poverty alleviation. This frontline includes many professionals working directly with poor people throughout the developing world, delivering health, education and other services aimed at helping people to help themselves. Our study focuses on a sample of 296 young Ethiopians who, at the time of the experiments reported here, were about to join this frontline as government-employed nurses and doctors. We know from previous research (Lindelow and Serneels 2005) that the institutional environment that these youngsters are about to enter is problematic in many regards: their pay will be low; the way in which most of them are to be allocated jobs by the government will take no account of their preferences, while the allocation process is circumvented and compromised by some who use network connections to secure their job of choice; they will be working alongside colleagues who are regularly absent, possibly because they are maintaining portfolios of jobs, and who expropriate public resources for private use; or they will be working unsupported and in isolation in remote postings; and their performance, whether it be good or bad, will rarely, if ever, be assessed or appropriately rewarded.

Within this context, intrinsic motivations could play a significant role in determining the performance of these youngsters, not least of all, by reducing their corruptibility. However, their intrinsic motivations might also be subject to erosion or compromise as they become aware of the reality of their situation. To explore these issues, our research will ultimately make use of data collected during repeated interactions with this sample of 296 new nurses and doctors. To date, there has been only one interaction, during which the youngsters took part in two experimental games and, then, completed tests of their knowledge

of medical theory and practice and a questionnaire about their family background, education, current training, paid work experience in the health sector to date, job preferences, and whether and to what extent they expect to become involved in some of the behaviours listed above. The function of the experimental games was to generate two separate proxies for the youngsters' intrinsic motivations, one relating to the value they place on equitable divisions of resources and one relating to their tendency to pass on resources that have been entrusted to them.

3. Experimental Design

The TPPG involves three players, A, B, and C. Play is anonymous and one-shot. A is given an initial endowment of money and invited offer some portion (between zero and 100 percent) of that endowment to B. C is given an initial endowment some of which he can choose to spend on punishing A. In Ethiopia, A's initial endowment was set at Birr 50 (approximately $5) and he or she could make allocations in multiples of Birr 5. C's initial endowment was set at Birr 25 and he or she could spend Birr 5 to reduce A's payoff by Birr 15 or Birr 10 to reduce A's payoff by Birr 30.[3] As made decisions by dividing their initial endowment between two envelopes. Cs had to fill out forms stating what they would do in the case of each possible division that could be made by A, before seeing the actual division made. Thus, we elicited eleven data points for each C.

The GTG is similar to Berg, Dickhaut and McCabe's (1995) investment game (referred to as the BDM below), in which a Proposer decides how much (zero to 100 percent) of her initial cash endowment to send to a Responder. The amount she sends is tripled by the experimenter before being passed on to the Responder who receives the same initial endowment. Then, the Responder is invited to send some portion (zero to 100 percent) of the tripled amount back to the Proposer. The GTG differs from this game in one respect: the Responders send money back, not to the Proposer who sent to them, but to some other Proposer present in the same session. In Ethiopia, the Proposers and Responders' initial endowments were set at Birr 40 and each player could send multiples of Birr 10. Proposers made their decision by dividing their initial endowment between two envelopes. Responders had to fill out a form stating what they would do in the case of each possible amount sent by their Proposer, before seeing the actual amount sent. Thus, we elicited four data points for each Responder.

We chose the TPPG because we wished to measure our survey respondents' valuations of what might be described as 'norms of fairness'. This being the case, we were particularly interested in how our 296

[3] In Fehr and Fischbacher (2004) original design the Player Cs could buy any amount of punishment, paying 1 unit for every 3 units of fine. We reduced the number of options available to the Player Cs in order to simplify training and minimize problems of cognition. The scripts and forms used during the experimental sessions are presented in Appendix B.

survey respondents would behave in role C. So, we recruited an additional 618 nursing and medical students to assume the roles A and B.[4] We chose the GTG because we wished to measure our survey respondents' tendencies to pass on resources that had been entrusted to them. In contrast to Responders in the more commonly played BDM investment game, the decision of Responders in the GTG is conceptually closer to the decision made on a day-to-day basis by public servants; they receive resources, which they are expected to pass on to others. So, here, we wanted each of our 296 survey respondents to assume the role of Responder. Hence, we retained half of the additional 618 subjects recruited for the TPPG to assume the Proposer role in the GTG.

Each of 927 subjects attended an experimental session held at the college in which they study. Students from eleven colleges were involved. One or two sessions were held in each college. Where two sessions were held, the sessions were run back-to-back and steps were taken to ensure that participants in the first session could not inform participants in the second about the games. A total of 20 sessions were held, with between 29 and 60 subjects involved in each. At the start of every session, the survey respondents (Player Cs) were shown into one classroom and the other students (Player As and Bs) into a second classroom. In each classroom, the students were sat at amply spaced desks. No talking was allowed. The TPPG was explained verbally and each subject was provided with a printed table showing what their earnings would be under all possible scenarios. The As each received two envelopes one containing their endowment and one into which they could place the amount of money they wished to pass on to B. The Player Bs received envelopes containing slips of paper informing them of their role and explaining that they would receive their earnings and the end of the session. The Cs, i.e., our survey respondents, were taught the game verbally, received the printed tables of earnings under all possible scenarios, and then asked to detail their strategy on the form provided. The envelopes and forms were processed at a desk in the front of the first classroom.

Before knowing the outcome of the TPPG, the students were taught the GTG. Again, the instructions were given verbally and each subject was supplied with a printed table of earnings under all possible scenarios. In the second classroom those students who had been passive Bs in the TPPG, then, received two envelopes, one containing a slip of paper informing them that they were Proposers and their initial endowment of Birr 40, the other to be used to send whatever amount they wished to a Responder. The remaining additional subjects received envelopes containing slips of paper informing them that they had not been picked to participate in the GTG but would receive compensation for being asked to sit quietly until the end of the experimental session. In the first classroom, the Responders were, again, asked to detail their strategy on the form provided.

[4] For these additional 618 students we have neither proxies for cognitive ability nor survey data relating to their personal characteristics.

All 20 sessions were led by a graduate student from Oxford University with the assistance of two Ethiopian research assistants. One research assistant was assigned to each room, while the Oxford graduate student remained in the first classroom throughout. The instructions were delivered verbally in Amharic by the research assistants using a script developed in English and then translated and back translated by professional translators who were not, otherwise, involved in the project.

Preliminary results indicate that the students' decisions within the games are correlated with their expected own future involvement in potentially corrupt or illegal behaviours. Those imposing greater fines on As who make low allocations to Bs in the TPPG and who pass on more in the GTG have lower expectations of becoming involved in bribery aimed at securing preferred public sector positions or moonlighting while holding such positions (Barr, Lindelow, Garcia-Montalvo, and Serneels, 2005).

4. Respondent characteristics

The characteristics of the 296 survey respondents for whom we have strategy data from the experiments are presented in Table 1. Their average age was 23 years. 40 percent were women. 91 percent had attended government-run secondary schools. On average, they had about four months of paid work experience in the health sector. And the average estimated annual expenditure of the households in which they grew up was just over 50,000Birr (approximately 6,000 USD).

Table 1: Characteristics of medical and nursing students

	Means and proportions			Sig. diff. between
	All	Medical	Nursing	sub-samples
Number of observations[+]	296	90	206	
doctor (proportion)	0.304			
age (years)	22.698	23.344	22.326	***
female (proportion)	0.402	0.144	0.507	***
private schooling (proportion)	0.090	0.133	0.080	*
experience in health sector (years)	0.276	0.041	0.356	***
estimated parental household expenditure (per year in 10,000s of Birr)	5.127	5.093	5.153	

Notes: significant difference in means between medical and nursing students - *** at 1% level, * at 10% level, + numbers of available observations fall to 295 in the case of age, 289 in the case of private schooling, and 292 in the case of estimated parental household expenditure.

We use the respondents' course of study as one of our proxies for their cognitive ability. Just over 30 percent of the respondents were medical students, while the rest were training to be nurses. We assume that a medical training is more cognitively demanding than a nurses training. However, note that the medical students are significantly older, have less experience working in the health sector, and are more likely to be men and to have attended a private secondary school. Thus, it will be important to control for these other characteristics during our analysis.

Our second proxy for cognitive abilities is the score that each respondent achieved in the tests of their knowledge of medical theory and practice. The tests were designed by a team of medical experts involved in teaching nurses and medical students in Ethiopia. This ensured that the tests took both the curriculum and Ethiopian conditions into account. Nursing and medical students sat different tests, although they followed a common structure. The tests contained two parts: (i) knowledge of evidence based medicine; and (ii) technical knowledge on how to deliver medical interventions. In total the nursing and medical students had to answer 53 and 52 multiple choice questions respectively. Each correct answer was given the same weight, and the final score was the sum of the correct replies. Both nursing and medical students scored better on the second part. Scores varied across schools, especially between nursing schools, possibly because different schools followed different curricula or attracted students of varying abilities.

These tests were designed to provide measures of the students' knowledge of medicine to be used in various analyses relating to their work performance rather than measures of their cognitive abilities in general. So, our analysis here is based on the assumption that the scores also reflect the students' innate and enhanced (by their studies) cognitive abilities. Because the nursing and medical students sat different tests, we can conduct only within course comparisons using the test scores. For the purposes of analysis we normalize the test scores for the medical and nursing students separately. We do not present descriptive statistics as within courses the mean test scores and their standard deviations equal one by construction and it is inappropriate to pool.

5. Strategy choice and cognition in the Third Party Punishment game

Fehr and Fischbacher (2004) found that in role C Swiss students generally chose fining strategies that declined monotonically as As' allocations to Bs increased from zero to 40 percent and indicated no fining of allocations from 50 to 100 percent. Such strategies can be explained by Fehr and Schmidt's (1999) model, which predicts that some individuals will be willing to pay to reduce the difference between their own and others' payoffs. They can also be explained by the models of Levine (1998) and Falk and Fischbacher (1999), which predict that some individuals will be willing to pay to punish others who act

unfairly, spitefully, or selfishly towards others, combined with the hypothesis made by Fehr and Fischbacher (2004) that the distribution norm that appears most salient to subjects in an experiment such as this is for A to allocate 50 percent of the stake to B.

Figures 1 and 2 summarize the fining strategies chosen by the Ethiopian survey respondents in the TPPG.[5] In Figure 1 the dark and lightly shaded portions of each bar indicate the frequencies with which the corresponding allocation by A to B attracted a high or a low fine respectively. Figure 2 shows the average expenditure on fining associated with each possible allocation. Allocations of less than 50 percent attracted the most and the highest fines and, in this range, lower allocations attracted higher average fines. Fewer, but still a significant (at the 1 percent level) proportion of respondents chose to fine allocations of more than 50 percent. In this range, allocations closer to 50 percent attracted more fines. Finally, a significant (at the 1 percent level) proportion of subjects chose to fine in the case of a 50:50 split.

Figure 1: Frequencies of high and low fines in the TPPG:

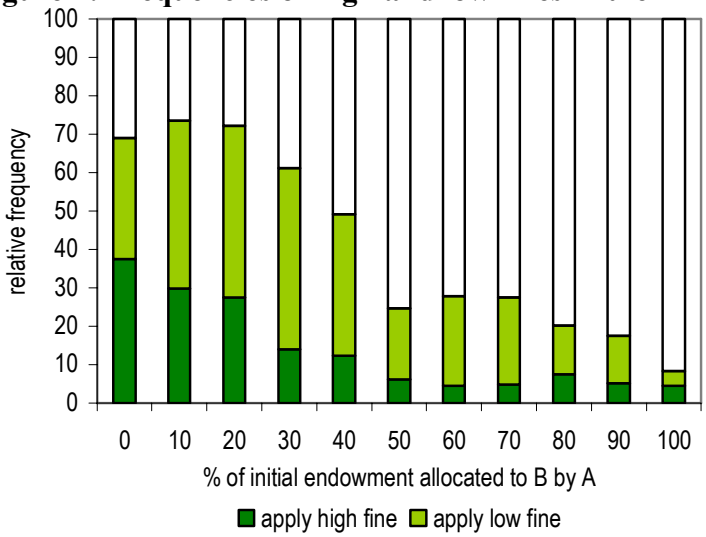
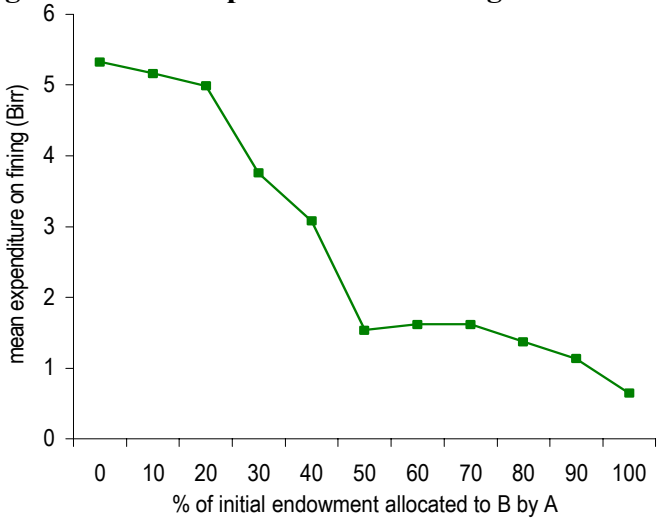


Figure 2: Mean expenditures on fining in the TPPG



[5] The behaviour of the Player As is presented in Appendix A.

These figures indicate that, while the strategies chosen by Ethiopian students are qualitatively similar to those chosen by the Swiss students in the zero to 40 percent range, at least some of them differ markedly in the 50 to 100 percent range. Significant levels of fining in the 50 to 100 percent range cannot be explained with reference to Fehr and Schmidt's (1999) model. However, the models of Levine (1998) and Falk and Fischbacher (1999) can explain fining in the 60 to 100 percent range if over-generosity is perceived as a selfish act, possibly because it is motivated by pride. They can also explain fines associated with the 50 percent allocation if combined with distribution norms advocating generosity in excess of equal sharing in the case of windfall gains. Below, we classify the Ethiopian students' strategies into four types, one consistent with the findings of Fehr and Fischbacher (2004), two relating to these alternative assumptions about which allocations are perceived as selfish, spiteful, or unfair, and one other capturing all the apparently random and internally inconsistent strategies. We, then explore the extent to which each type is associated with problems of cognition. The four strategy-types are listed and described in Table 2.

Table 2: Types and frequencies of fining strategy

Proportion (%) of fining strategies that are…	Medical 90	Nursing 206	All 296
FF strategies: monotonically declining (or zero) in 0 to 45% range, zero in 50 to 100% range	65.56%	42.72%	49.66%
U strategies: U-shaped and zero at 50%	1.11%	0.00%	0.34%
WG strategies: monotonically declining in 0 to 100% range and not zero at 50%	13.33%	9.71%	10.81%
R strategies: none of the above, appear random and internally inconsistent	20.00%	47.09%	38.85%
zero throughout 0 to 100% range	8.89%	15.05%	13.18%

Notes: Chi test indicates that distributions across strategies vary significantly (1% level) between medical and nursing students. Nursing students are significantly more likely to pick a strategy that is neither monotonically declining or U-shaped.

The first strategy-type includes strategies specifying monotonically declining fines as As' allocations to Bs increase from zero to 40 percent and no fines relating to allocations of 50 percent and above. It also includes strategies indicating no fining. Below, we refer to this strategy-type as FF in recognition of the fact that they are similar to those observed by Fehr and Fischbacher (2004). The second strategy-type specifies monotonically declining fines as As' allocations to Bs increase from zero to 40 percent, no fining in the case of a 50 percent allocation, and monotonically increasing fines as allocations increase from 60 to 100 percent. Below, we refer to this strategy-type as U (with reference to its shape). The third strategy-type specifies monotonically declining fines as allocations increase from zero to 100 percent and a positive level of fining when the allocation is 50 percent. Below, we refer to this strategy-type as WG (for windfall gain). The fourth strategy-type includes all apparently random and internally inconsistent strategies. Below, we refer to this strategy-type as R (for random).

Approximately half of the Ethiopian students chose FF strategies, with 13 percent choosing the selfish money-maximizing strategy, only one chose a U strategy, over 10 percent of chose WG strategies, and nearly 40 percent specified R strategies.

Table 2 also presents the proportions of medical and nursing students that chose each strategy-type and thereby provides a basis for our first test of a relationship between strategy choice and cognitive ability. A Chi-test indicates that choice of strategy-type differs significantly (1 percent level) between medical and nursing students. Among the more cognitively able medical students, 66 percent chose FF strategies, while only 20 percent chose R strategies. In contrast, only 43 percent of nursing students chose FF strategies, while 47 percent chose R strategies. WG strategies were chosen by statistically similar proportions of medical and nursing students (13 and 9 percent respectively). The one student choosing a U strategy was studying medicine. These preliminary results suggest that the less cognitively able are more likely to specify apparently random and internally inconsistent strategies. While it is not possible to draw any conclusions based on the one student who chose a U strategy, the choice of a WG strategy does not appear to be associated with problems of cognition.

Table 3: Types of fining strategies and normalized test scores

Normalized test scores associated with fining strategies that are…	Medical 90	Nursing 206
FF strategies	0.232	0.011
U strategies	-0.010	
WG strategies	-0.184	-0.052
R strategies	-0.639 ***	0.001

Notes: *** significantly different to scores of those chosing FF strategy at 1% level according to a Wilcoxon rank-sum (Mann-Whitney) test.

In Table 3 we present the average normalized tests scores of the medical and nursing students choosing each of the four strategy-types. Thus, we see that the medical students choosing R strategies had significantly (at the one percent level according to a Wilcoxon (Mann-Whitney) rank sum test) lower test scores than those choosing FF strategies, whereas those choosing WG strategies did not. Among the nursing students, tests scores did not predict strategy choice.

In Table 4 we present a series of multinominal logit regressions based on the sample of students choosing FF, WG or R strategies. We omit the one student who chose a U strategy. In each regression, the basis for comparison is those students choosing FF strategies. However, in the text below we also present the

results of tests, based on these regressions, that compare those choosing WG and R strategies. In all the regressions the standard errors are adjusted to control for possible non-independence within sessions.

Table 4: Regression analysis of fining strategy characteristics

	Type of strategy chosen (multinomial logit estimations)			
	Medical	Nursing	All	All
WG strategy (compared to FF strategy)				
constant	-1.579 ***	-1.483 ***	-1.483 ***	-6.056 ***
	(0.303)	(0.258)	(0.255)	(1.902)
doctor (dummy)			-0.096	-0.093
			(0.382)	(0.472)
normalized test score	-0.494	-0.063	-0.063	-0.043
	(0.328)	(0.150)	(0.148)	(0.179)
normalized test score x doctor			-0.431	-0.425
			(0.342)	(0.395)
age (years)				0.099
				(0.074)
female (dummy)				-0.024
				(0.248)
private schooling (dummy)				0.816
				(0.750)
experience in health sector (years)				-0.261
				(0.235)
estimated parental household expenditure				0.415
				(0.331)
R strategy (compared to FF strategy)				
constant	-1.363 ***	0.108	0.108	-0.858
	(0.207)	(0.207)	(0.204)	(1.692)
doctor (dummy)			-1.470 ***	-1.246 ***
			(0.281)	(0.350)
normalized test score	-0.962 **	-0.010	-0.010	0.054
	(0.488)	(0.138)	(0.137)	(0.166)
normalized test score x doctor			-0.952 **	-1.006 **
			(0.477)	(0.496)
age (years)				-0.027
				(0.053)
female (dummy)				0.290
				(0.318)
private schooling (dummy)				0.229
				(0.648)
experience in health sector (years)				0.244
				(0.074)
estimated parental household expenditure				0.251
				(0.249)
Observations	89	206	295	286
Pseudo Rsq.	0.0736	0.0002	0.0571	0.0836

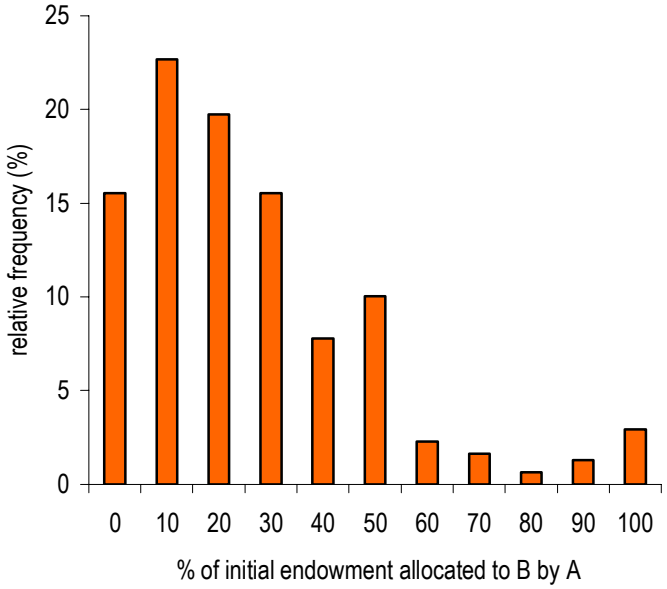
Notes: *** significant at the 1% level, ** significant at the 5% level, * significant at the 10% level.

In the first two columns we regress choice of strategy-type on the normalized test scores of medical and nursing students respectively. These regressions confirm the results of the Chi and Wilcoxon (Mann-Whitney) rank sum tests reported above. Medical students are significantly less likely to choose WG and R strategies as compared to FF strategies and, among the medical students, the less cognitively able are more likely to choose R strategies. Nursing students are less likely to choose WG strategies than FF

strategies, but are similarly likely to choose FF or R strategies and appear not to sort across strategy-types by test score. These results are also confirmed by the regression in the third column, which is based on the pooled sample of medical and nursing students and contains as regressors a dummy variable identifying the medical students, the normalized test scores and the interaction between these two variables. In the fourth column we add the students' ages, sex, school-type, experience in the health sector, and estimated parental household expenditures as control variables. The previously reported findings remain, while none of the additional control variables are significant.

Tests of linear restrictions applied to the models presented in Table 4 also indicate that medical students are significantly more likely to choose WG strategies than R strategies and that, among medical students, the more cognitively able are more likely to choose WG strategies than R strategies. Such tests also indicate no sorting of nursing students across strategy-types (WG and R in this case) by test score.

Figure 3: Allocations to Bs by As in the TPPG



To summarize, our analysis indicates that subjects choosing apparently random and internally inconsistent strategies are distinct from subjects choosing all other types of strategy. According to both of our proxy measures they are less cognitively able. Thus, we can conclude that such strategies are symptomatic of problems of subject understanding. In contrast, there is no evidence that the significant proportion of students choosing monotonically declining strategies indicating positive fines in the case of equal sharing were suffering such problems. This being the case, we should consider the alternative explanation, i.e., that those students differ from both their own colleagues and the Swiss students in Fehr and Fischbacher's (2004) original experiment with respect to the behaviour they perceive as being punishable. Their fining strategies indicate that they believe As should pass more than 50 percent of their

original endowment on to Bs. That As' original endowment is a windfall gain may be salient here. While the behavior of the Ethiopian As in the TPPG is of less relevance to our current endeavour, it is worth noting that, consistent with the latter strategy-type, nine percent (significant at the one percent level) of the As made allocations of between 50 and 100 percent of the initial stake (see figure 3).

6. Strategy choice and cognition in the Generalized Trust Game

In the GTG, selfish money-maximizing Responders will return nothing regardless of what they have been sent. Several models, including Fehr and Schmidt's (1999), Levine's (1998) and Falk and Fischbacher's (1999), predict that, in the standard BDM, Responders will return more the more they receive. However, in the GTG the Responders do not know how much the Proposers to which they are returning money sent. Hence, they do not know how inequitable the resource distribution between themselves and the Proposers to which they are sending is and have no information upon which to base judgements about their preferences or actions. Theories of generalized reciprocity (e.g., Takahashi (2000) and Bearman (1997)) predict that received resources will be passed on. However, when applied to the GTG, it is unclear whether this prediction relates to the amount received before or after it has been tripled and, hence, leaves us none the wiser as to what type of strategy to look for.

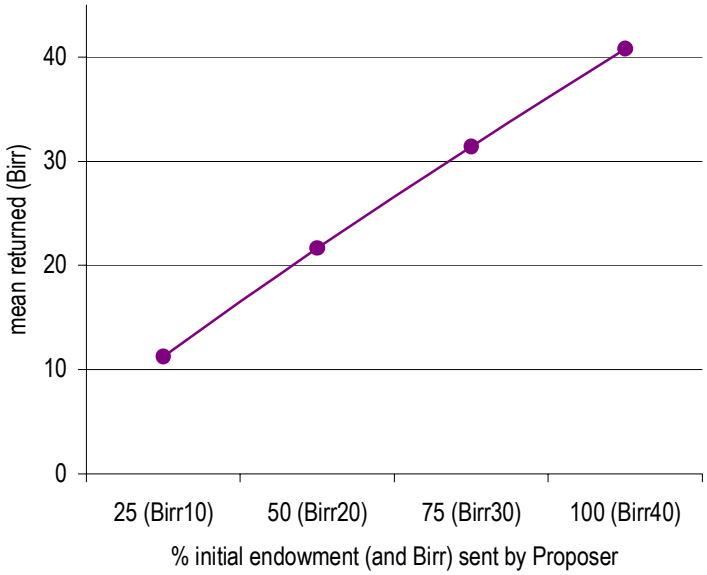
The objective of Buchan, Croson, and Dawes' (2002) study was not to test theoretical predictions about strategy characteristics. Their aim was to compare generalized to pairwise trusting and reciprocating behaviour. However, during the analysis of their Responder data, generated using a sequential decision protocol, they made an implicit assumption about latent strategy form. When comparing behaviour across subjects, they focused on the ratio of the amount returned by the Responder to the amount sent by the Proposer.[6] The comparability of this ratio across subjects depends on the assumption that the subjects' latent reciprocating strategies are rays, i.e., they are linear in the amount received from Proposers and pass through the (nothing sent, nothing returned) origin. This assumption is clearly of practical value, but is it correct? Our objective in this section of the paper is to test this assumption by establishing whether deviations from such ray-type strategies are errors associated with problems in subject cognition.

Figure 4 plots the mean amounts that Responders would return given each possible amount (prior to tripling) received from a Proposer. The figure indicates that, on average, Responders would send back marginally more than they receive. Based on a subject fixed effects regression of amounts returned against amounts received, we cannot reject the hypothesis that the slope of this graph is equal to one,

[6] Elsewhere, Croson and Buchan (1999, 2003) apply this approach to sequential decision data from Responders in the BDM game. Other researchers, including Barr (2003), Burns (2004), and Asraf, Bohnet and Piakov (2003), followed a similar approach.

although the hypothesis that the graph passes through the origin is rejected at the 10 percent level, even though the intercept is less than 2 Birr. These results provide some support for the models of generalized reciprocity cited above and indicate that the amount sent by the original Proposer and not the tripled amount is salient. However, the positive intercept is not consistent with the assumption that the subject's strategies are rays. Further, the regression explains only 22 percent of the variation in behaviour, indicating that there is considerable variation across subjects.

Figure 4: Reciprocating strategies in the GTG: Mean return associated with amounts sent



In Table 5 we classify the Responders' strategies into four types. The first strategy-type is linear and passes through the origin. This includes the selfish money-maximizing strategy involving no return ever and strategies that specify that the amount sent, twice the amount sent, or three times the amount sent would be returned. We refer to these as RAY strategies. The second strategy-type includes other monotonically increasing strategies. We refer to these as MON (for monotonically increasing) strategies. It is possible that these strategies were chosen by Responders who would have specified a ray with a slope other than zero, one, two or three had their possible response set not been restricted to multiples of five Birr. We return to this issue below. The third strategy-type is motivated by Greig and Bohnet (2005) who found that a large proportion of their Kenyan Responders in a BDM game chose to return the same small amount regardless of how much they received. We refer to such strategies as FLAT (referring to the shape of the strategy). These are not consistent with the assumption that all subjects have latent ray-type strategies. Our final strategy-type encompasses strategies that decline over all or part of the range of possible amounts received. We refer to these as OTH (for other) strategies. These, too, are not consistent with the assumption that subjects have latent ray-type strategies.

In Ethiopia, a third of the students specified a RAY strategy: 8 percent specified the selfish money-maximizing strategy, 16 percent chose to return the amount they received prior to it being tripled, seven percent chose to return double, and two percent chose to return the tripled amount. A further 42 percent of the students specified MON strategies, only 3 students chose a FLAT strategy, and 23 percent chose OTH strategies.

Table 5: Types of reciprocating strategies

Proportion (%) of reciprocating strategies that are…	Medical 90	Nursing 206	All 296
RAY strategies: linear and passing through (0,0)	40.00%	30.10%	33.11%
zero throughout	*5.56%*	*9.71%*	*8.45%*
linear with a slope of 1	*22.22%*	*12.62%*	*15.54%*
linear with a slope of 2	*12.22%*	*4.85%*	*7.09%*
linear with a slope of 3	*0.00%*	*2.91%*	*2.03%*
MON strategies: monotonically increasing but not linear	51.11%	39.32%	42.91%
FLAT strategies: flat at 10 Birr throughout	1.11%	0.97%	1.01%
OTH strategies: neither monotonically increasing nor flat	7.78%	29.61%	22.97%

Notes: Chi test indicates that distributions across strategies vary significantly (1% level) between medical and nursing students. Nursing students are significantly more likely to pick a strategy that is not monotonically increasing.

Table 5 also presents the proportions of medical and nursing students that chose each strategy-type and thereby provides a basis for our first test of a relationship between strategy choice in the GTG and cognitive abilities. A Chi-test indicates that choice of strategy-type differs significantly (1 percent level) between medical and nursing students. Only 8 percent of the more cognitively able medical students chose OTH strategies, while 30 percent of nursing students chose such strategies. The division of the remaining strategies between RAYs and MONs is indistinguishable between students on the two courses of study. These preliminary results suggest that strategy choices that are not consistent with the assumption that all subjects have latent ray-type strategies may be associated with problems of cognition.

Table 6: Types of reciprocating strategies and normalized test scores

Normalized test scores associated with reciprocating strategies that are…	Medical 90	Nursing 206
RAY strategies	-0.026	-0.048
MON strategies	0.003	0.028
FLAT strategies	-0.904	
OTH strategies	0.245	0.040

Notes: No significant differences in test scores across groups by strategy-type.

In Table 6 we present the average normalized test scores of the medical and nursing students choosing each of the four strategy-types. Wilcoxon rank-sum (Mann-Whitney) tests indicate that among neither medical nor nursing students do test scores predict strategy-type choice.

Table 7: Regression analysis of reciprocating strategy characteristics

	Type of strategy chosen (multinomial logit estimations)			
	Medical	Nursing	All	All
MON strategy (compared to RAY strategy)				
constant	0.245	0.268	0.268	-1.411
	(0.208)	(0.249)	(0.246)	(1.657)
doctor (dummy)			-0.023	0.142
			(0.314)	(0.364)
normalized test score	0.030	0.075	0.075	0.132
	(0.129)	(0.119)	(0.117)	(0.144)
normalized test score x doctor			-0.046	-0.066
			(0.168)	(0.177)
age (years)				0.040
				(0.055)
female (dummy)				0.366
				(0.354)
private schooling (dummy)				-0.352
				(0.595)
experience in health sector (years)				0.103
				(0.138)
estimated parental household expenditure				0.111
				(0.229)
OTH strategy (compared to RAY strategy)				
constant	-1.670 **	-0.016	-0.016	-0.113
	(0.477)	(0.293)	(0.290)	(1.906)
doctor (dummy)			-1.654 ***	-1.487 **
			(0.532)	(0.597)
normalized test score	0.288	0.088	0.088	0.214
	(0.385)	(0.154)	(0.153)	(0.162)
normalized test score x doctor			0.200	0.062
			(0.392)	(0.410)
age (years)				0.059
				(0.065)
female (dummy)				0.681
				(0.570)
private schooling (dummy)				-0.257
				(0.421)
experience in health sector (years)				$2.6e^{-5}$
				(0.140)
estimated parental household expenditure				-0.305
				(0.237)
Observations	89	204	293	283
Pseudo Rsq.	0.0028	0.0006	0.0325	0.0485

Notes: ** significant at the 5% level, * significant at the 10% level.

In Table 7 we present a series of multinominal logit regressions based on the sample of students choosing RAY, MON or OTH strategies. We omit the three students who chose FLAT strategies. In each regression, the basis of comparison is those choosing RAY strategies. In all the regressions the standard errors are adjusted to control for possible non-independence within sessions. In the first two columns we

regress choice of strategy-type on the normalized test scores of medical and nursing students respectively. These regressions confirm that medical students are significantly less likely to choose OTH strategies than RAY strategies and that test scores do not predict strategy-type. These results are also confirmed by the regression in the third column, which is based on the pooled sample of medical and nursing students and contains as regressors a dummy variable identifying the medical students, the normalized test scores and the interaction between these two variables. In addition, a linear restriction test based on this regression indicates that medical students are also significantly less likely to choose OTH strategies than MON strategies. In the fourth column we add the students' ages, sex, school-type, experience in the health sector, and estimated parental household expenditure as control variables. The previously reported findings remain unchanged and none of the additional control variables are significant.

Recall that our objective in this section is to establish whether deviations from ray-type strategies are due to subject error or misunderstanding. The finding that more cognitively able medical students are more likely to choose RAY strategies than the less cognitively able nursing students indicates that this might indeed be the case. That those students who chose MON strategies are no less cognitively able than those who chose RAY strategies might also support this hypothesis, but only if the former represent frustrated attempts to specify a ray with a slope other than zero, one, two or three. In an attempt to investigate this issue further, we conduct one final analysis focused specifically on MON strategies. The aim of the analysis is to establish whether, within this sub-sample of students, those who are more cognitively able choose strategies that more closely approximate ray-type strategies.

In order to do this we, first, fit a ray to each of the sub-sample of students' strategies, i.e., we regress the amount they would return given each of the amounts they could receive on those received amounts, while specifying a zero intercept.[7] Then, taking the standard errors associated with these estimated coefficients as a proxy for how far each MON strategy deviates from a ray, we investigate the relationship between these deviations and the students' cognitive abilities. There is one complicating factor: the estimated slope coefficients and associated standard errors are positively correlated; possibly, those students with steeper latent strategies were more constrained by the need to specify their returns in multiples of five. Thus, we need to control for the estimated slope of the ray during our analysis.

The results of this exercise are presented in Table 8. The OLS regression in the first column simply demonstrates the positive correlation between the standard error and the estimated slope of the ray. In the second column we add the dummy variable that identifies medical students. Thus, we see that medical students pick strategies that more closely approximate rays than the less cognitively able nursing students.

[7] The resulting slope coefficients vary between 0.133 and 2.800. And of these coefficients, 85 percent are significantly different from zero at the one percent level and only two percent are insignificant at either the five or ten percent level.

In the third and fourth columns we introduce the normalized test scores, while focusing on the sub-samples of medical and nursing students respectively. Here, we find that, among medical students, the strategies chosen by the more cognitively able more closely approximate rays.

Table 8: Regression analysis of MON strategies

	Deviation from a pure RAY strategy (OLS estimations)			
	All	All	Medical	Nursing
constant	0.060 ***	0.065 ***	0.095 ***	0.062 ***
	(0.015)	(0.015)	(0.021)	(0.018)
estimated slope of ray	0.044 ***	0.047 ***	0.005	0.046 ***
	(0.011)	(0.011)	(0.015)	(0.014)
doctor (dummy)		-0.025 **		
		(0.012)		
normalized test score			-0.013 *	-0.002
			(0.008)	(0.008)
Observations	131	131	46	81
Rsq.	0.107	0.134	0.068	0.115

Notes: *** significant at 1% level, ** significant at 5% level, * significant at 10% level

7. Summary and conclusions

Our aim in this paper was to explore and demonstrate the value of the strategy method to field experimentalists. Specifically, we wished to demonstrate that, while the application of strategy method may lead to reductions in subject understanding, it may also generate insights, valuable not only within the context of the study in question but also to researchers planning new studies. Guided by the context of our study, we played two games, the TPPG and the GTG, with a fairly large sample of medical and nursing students, applying the strategy method to the role of particular interest in each case. Then, making use of two proxy measures for the students' cognitive abilities, we investigated the relationship between strategy-type choices and student understanding in each game.

Thus, in the TPPG, we found support for the assertion made by Bahry and Wilson (2003), that apparently random and internally inconsistent strategies are symptomatic of problems of cognition. In the absence of strategy data, it would not have been possible to test this assertion. During the same analysis, we also found that an observed difference in third party punishing behavior between Swiss and Ethiopian students is due, not to problems of cognition, but to variations in what is perceive as punishable behavior. A significant proportion (ten percent) of the Ethiopian students chose strategies indicating that they considered it appropriate for students in role A to allocate more than 50 percent of their initial endowment to B. And, indeed, a significant proportion of students in role A did precisely that. This is consistent with the existence of a norm advocating generosity in excess of equal sharing in the case of windfall gains.

Had we not adopted the strategy method, it is doubtful that we would have been able to challenge Fehr and Fischbacher's (2004) assumption about the salient distribution norm in this way.

Applying similar methods to the strategies of Responders in the GTG, we found evidence supporting an assumption often but not explicitly made by researchers comparing reciprocating behavior across subjects; that deviations from linear strategies passing through the (nothing sent, nothing returned) origin are symptomatic of subject error. Again, without strategy data, it would have been impossible to test this assumption.

So, should field experimentalists apply the strategy method wherever possible? On reflection, we believe not. In the TPPG, nearly 40 percent of our subjects chose strategies symptomatic of problems of understanding. This proportion is large, especially as, in contrast to many field subjects, ours are literate and familiar with abstract problem solving. If we assume that the game itself is fairly easy to understand, these problems could be due to either our decision to apply the strategy method or our decisions about how to apply it. Given the success of the Cross Cultural group in applying the strategy method in TPPGs played with subjects who are neither literate nor used to abstract problem solving (papers forthcoming), we suspect the latter. We required our subjects to indicate their chosen strategies on forms through a series of multiple choice questions. Whereas, in the Cross Cultural project, strategies were elicited verbally with the aid of props (real money moved around on a tabletop divided into areas relating to each player) only after the individual subjects had been taught the game one-on-one and their understanding had been thoroughly tested. If this is the reason for the problems in our data, field experimentalists contemplating the TPPG will have to decide how much time and effort they are willing and able to invest in the quality of strategy data. If resources are limited, sequential decision making may be advisable. However, then data quality will be unknown, cross-subject comparability will be compromised, and far fewer insights into subject motivations will be gained.

In the GTG, only 23 percent of the Ethiopian students chose strategies symptomatic of problems of understanding. However, our analysis suggests that some of the 42 percent who chose non-linear, monotonically increasing strategies may also have had some problems understanding the game or the related maths. Once again, investing more time in subject training and testing could considerably reduce the incidence of such problems. However, in this case, the application of the sequential decision making method may have fewer disadvantages: data quality would be unknown, but, given our result relating to deviations from ray-type strategies, few insights relating to subject motivations will be lost, and subject comparisons based on the ratio of what they return to what they receive are likely to be well founded.

References

Ashraf, N, I. Bohnet and N. Piankov (2003): "Is trust a bad investment?", Faculty research working paper, J.F. Kennedy School of Government, Harvard university.

Bahry, and R. Wilson (2003) 'Confusion or fairness in the field? Rejections in the Ultimatum Game under the strategy method' presented at a conference on field experiments, Middlebury College, April, 2003.

Barr, A. 'Trust and Expected Trustworthiness: Experimental evidence from Zimbabwean villages,' *Economic Journal*, 113(489): 614-630.

Barr, A., Garcia-Montalvo J. Lindelow M. and P. Serneels (2005) 'Intrinsic motivations on the development frontline: Do they exist? Do they endure?' GPRG WPS/2005-?

Bearman, P. (1997) 'Generalized exchange' *American Journal of Sociology,* 102:1383-1415.

Berg, J., J. Dickhaut, and K. McCabe, K. (1995) 'Trust, Reciprocity, and Social History' *Games and Economic Behavior*, 10: 122-42.

Blount, S. and M. Bazerman (1996) 'The inconsistent evaluation of absolute versus comparative payoffs in labor supply and bargainning' *Journal of Economic Behavior and Organization,* 30: 227-240.

Bolton, G. E., and A. Okenfels (2000) 'ERC: A theory of equity, reciprocity, and competition' *American Economic Review,* 90: 166-93.

Brandts, J. and G. Charness (2000) 'Hot vs. cold: Sequential responses and preference stability in experimental games' *Experimental Economics*, 2: 227-238.

Brosig, J., J. Weimann, and C.-L. Yang (2003) 'The hot versus cold effect in a simple bargaining experiment' *Experimental Economics,* 6(1): 75-90.

Buchan, N. R., R. T. A. Croson, R. M. Dawes (2002) 'Swift Neighbors and persistent strangers: A cross-cultural investigation of trust and reciprocity in social exchange' *American Journal of Sociology,* 108(1): 168-206.

Burns J. (2004): "Race and trust in post apartheid South Africa", mimeo, University of Cape Town and SFI.

Croson, R. and N. Buchan (1999), "Gender and culture: International experimental evidence from trust games", The American Economic Review 89 (2), 386-391

Croson, R. and N. Buchan (2003), "The boundaries of trust: Own and other's actions in US and China", *Journal of Economic Behavior and Organization*, forthcoming.

Falk, A. and U. Fischbacher (1999) 'A theory of reciprocity', CEPR Discussion Paper No. 3014 .

Fehr, E. and U. Fischbacher (2004) 'Third-party punishment and social norms' *Evolution and Human Behavior,* 25: 63-87.

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

Fischbacher, U., S. Gächter, and E. Fehr (2001) 'Are People Conditionally Cooperative? Evidence from a Public Goods Experiment' *Economics-Letters*. June 2001; 71(3): 397-404.

Greig. F., and I. Bohnet (2005) 'Is There Reciprocity in a Reciprocal Exchange Economy? Evidence from a Slum in Nairobi, Kenya' KSG Working Paper Number: RWP05-044.

Güth, W., S. Huck, and W. Mueller (2001) 'The relevance of equal splits in ultimatum games' *Games and Economic Behavior,* 37: 161-169.

Henning-Schmidt, H., Zhuyu L. and Chialiang Y (2001) 'Non-monotone strategies in ultimatum bargaining: First results from a video experiment in the People's Republic of China' Proceedings of the International Congress of Mathematicians Game Theory and Applications Satelite Conference

Levine, D. (1998) 'Modelling altruism and spitefulness in experiments' *Review of Economic Dynamics,* I: 593-622.

Lindelow, M. and P. Serneels (2005) 'The performance of health workers in Ethiopia: results from qualitative research' Social Science and Medicine, forthcoming.

McKelvey, R. D., and T. R. Palfrey (1998) 'Quantal Response Equilibria for Normal-Form Games' *Games and Economic Behaviour*, 10: 6-38.

McKelvey, R. D., and T. R. Palfrey (1998) 'Quantal Response Equilibria for Extensive-Form Games' *Experimental Economics*, 1(1): 9-41.

Oxoby, R. J., and K. N. McLeish (2004) 'Sequential decision and strategy vector methods in ultimatum bargaining: Evidence on the strength of other regarding behaviour' *Economic Letters* 84: 399-405.

Selton, R. (1967) 'Die strategiemethode zur erforschung des eingeschränkt rationalen verhaltens in rahmen eimes oligopolexperiments' in Sauerman, H. (ed.) *Beiträge zur Experimentellen Wirtschaftsforschung,* Tübingen: J.C.B. Mohr (Paul Siebeck).

Takahashi, N. (2000) 'The emergence of generalized exchange' *American Journal of Sociology,* 105: 1105-34.

World Bank (2003) *World Development Report 2004: Making Services Work for Poor People*, Washington DC: World Bank.

Appendix A: Behavior of Proposers in the GTG