CBESS Discussion Paper 09-17

A Test of the Experimental Method

**by Shaun P. Hargreaves Heap,
Arjan Verschoor, and Daniel John Zizzo**

CBESS, University of East Anglia

Abstract
Do the insights into human behavior generated by laboratory experiments hold outside the lab? This is the crucial question of external validity that naturally troubles both experimentalists and their critics. We address this question by adopting Popper's injunction that hypotheses should be tested, not by seeking instances of confirmation, but through exposure to conditions where falsification is a serious possibility. We select a population where the non-experimental evidence points to behavior which is quite unlike what is typically found in the laboratory and we examine whether their experimental results track these untypical behaviors. In our case, they do.

JEL classification codes
B41, C72, C93, Z13

Keywords
external validity, experiments, trust, groups, discrimination

CBESS
University of East Anglia
Norwich NR4 7TJ
United Kingdom
www.uea.ac.uk/ssf/cbess

A Test of the Experimental Method

Shaun P. Hargreaves Heap[1]
School of Economics and CBESS
University of East Anglia
Norwich NR4 7TJ, United Kindgom

Arjan Verschoor[2]
School of International Development and CBESS
University of East Anglia
Norwich NR4 7TJ, United Kingdom

Daniel John Zizzo[3]
School of Economics and CBESS
University of East Anglia
Norwich NR4 7TJ, United Kingdom

December 2009

Abstract

Do the insights into human behavior generated by laboratory experiments hold outside the lab? This is the crucial question of external validity that naturally troubles both experimentalists and their critics. We address this question by adopting Popper's injunction that hypotheses should be tested, not by seeking instances of confirmation, but through exposure to conditions where falsification is a serious possibility. We select a population where the non-experimental evidence points to behavior which is quite unlike what is typically found in the laboratory and we examine whether their experimental results track these untypical behaviors. In our case, they do.

Keywords: external validity, experiments, trust, groups, discrimination.
JEL Classification Codes: B41, C72, C93, Z13.

[1] Tel: +44-1603-593422; fax +44-1603-456259. Email address: s.hargreavesheap@uea.ac.uk
[2] Tel: +44-1603-592806; fax +44-1603-451999. Email address: a.verschoor@uea.ac.uk
[3] Tel: +44-1603-593668; fax: +44-1603-456259. Email address: d.zizzo@uea.ac.uk
We wish to thank our field research manager Matthew Osborne for research assistance, and the University of East Anglia for financial support. The usual disclaimer applies. The experimental instructions can be found at: http://www.uea.ac.uk/~ec601/trustdevinstr.pdf

Do insights into human behavior generated by laboratory experiments hold outside the lab? This is the question of external validity.[4] It troubles experimentalists and their critics alike and, with the burgeoning of experimental work in economics, the issue has assumed increasing importance. We address it here in a way that is distinct because we adopt Popper's injunction that hypotheses (like the one of external validity) should be tested, not by seeking instances of confirmation, but through exposure to conditions where falsification is a serious possibility.

Typically, there are two approaches to assessing the hypothesis that experiments have external validity. While both play important roles in the judgment over whether experimental results generalize to the world outside the lab, they are not without problems. One approach examines the consistency between the experimental results and the evidence from field studies/experiments and the econometric analysis of naturally occurring behaviors (see Francesco Guala, 2005 and Steven Levitt and John List, 2007). The other tests, in the laboratory, the robustness of the behavioral insights to changes in design that judiciously introduce features which capture the way that decision problems are encountered outside the laboratory (see Levitt and List, 2007). The latter is open to the objection that it appears to be premised on the external validity of experiments.[5] The first is potentially subject to the criticism that confirmations of this sort are weak tests of a theory, not least because the experimentalist has some discretion in selecting the non-experimental evidence to check against. As Karl Popper (1963) remarked, 'it is easy to obtain confirmations, or verifications, for nearly every theory…' (p.47)[6]. He went on to argue that 'confirming evidence should not count *except when it is the result of a genuine test of the theory*; and this means that it can be presented as a serious but unsuccessful attempt to falsify the theory' (p.48). More recently, Deborah G. Mayo (1996) and Mayo and Aris Spanos (2005), among others, have stressed the importance of *severe tests* of theories.[7]

This is what we aim to do. That is, we present a serious attempt to falsify and provide a severe test of the claim of external validity.

We select a population where the non-experimental evidence points to behavior which is quite unlike what is typically found in the laboratory and conduct an experiment with this population to see whether the experimental results track these 'unusual' behaviors. This is a 'risky' or 'serious' test in the Popperian sense because, if the behavior that is typically found in the laboratory is in some central way an artefact of the laboratory setting, then the experimental method is less likely in these circumstances to produce behavior that is consistent with the non-experimental evidence.

For this purpose, we focus on two behavioral regularities in the experimental literature. One is the presence of pro-social behaviors (see James Andreoni, William T. Harburgh and Lise Vesterlund, 2007). People often appear motivated by some social or other regarding preference: for example, when they contribute to a public good, decide to trust another, give a significant amount as a 'dictator', and so on. The other is that, in the presence of groups, people exhibit a form of

[4] Internal validity is concerned with whether the experiment has been designed in such a way that clear inferences may be drawn regarding behavior in the laboratory.

[5] This problem arises here in the same way that proofs of induction always seems to turn on accepting this principle in the first place because the evidence on how behaviour is sensitive to the kinds of change in context encountered outside the laboratory comes from laboratory experiments. Again like the so-called problem of induction, this problem while real is not a decisive argument against this approach.

[6] And of course, in support of Popper on this point, we might also add that the prevalence of a 'confirmation bias' is now well known in the experimental literature (see Rabin, 1998).

[7] While we are casting our argument in terms of Popper, it is useful to note that, as the references to Mayo (1995) and Mayp and Spanos (2005) suggest, it does not require a wider belief in a Popperian philosophy of science. Few methodologists would demur from Popper as a recommendation, in terms of the usefulness of severe tests such as the one we propose in this paper, though most would of course recognise that matters are more complicated than Popper (sometimes suggested) due to Duhem-Quine type unfalsifiability problems. Everyone would also agree that the practice of science is more complex, for example for the reasons highlighted by Thomas Kuhn (1970).

'ingroup' favoritism: that is, they trust fellow members more than outside, will contribute more to a public good with fellow group members and so on (see Gary Charness, Luca Rigotti, and Aldo Rustichini, 2007, and Shaun Hargreaves Heap and Daniel J. Zizzo, 2009). It is not difficult to find broad support for both experimental regularities in the non-experimental evidence in the manner of 'verificationism' (e.g., Robert D. Putnam, Robert Leonardi and Rafaella Y. Nanetti Putnam, 1993, Adalbert Mayer and Steven L. Puller, 2008, and Rashmi H. Assudani, 2009). We seek instead a population that appears on the basis of non-experimental evidence to be both especially individualistic in ways that make pro-social behavior unlikely and disinclined to 'ingroup' favoritism. The specific hypotheses, based on a belief in the external validity of the experimental method, are that an experiment on this population will a) reveal less pro-social behavior than is usual in experiments, and b) where pro-sociality does arise, insiders will not be favored.

The population that we selected for this test were the Gisu in Uganda. We discuss the non-experimental evidence that explains this choice in the next section. Section II describes the trust game experiment that we conducted with the Gisu and Section III reports the results and IV discusses them. We cannot reject both hypotheses. In other words, our experiment tracks the non-experimental evidence by producing similarly 'unusual' results.

I. The Gisu

> 'The Gisu, an agricultural people living on the slopes of Mount Elgon in eastern Uganda, have a long-standing reputation for quarrelsome and unruly behavior. Resort to violence in personal disputes is common among men and women. Conjugal disputes, for example, frequently involve physical injury to either or both spouses and the destruction of property. Beer parties are often the scenes of brawls, some of which may result in death.' (Jean La Fontaine, 1969, p.248)

While this quotation by an early and famous ethnographer may seem one-sided, by overlooking what one suspects are the many compensating attributes of the Gisu, it does suggest that they might be an appropriate subject pool for our experiment. With a reputation for 'quarrelsome and unruly behavior', it would be surprising under the assumption of external validity if they exhibited anything like the usual pro-social behaviors in a laboratory test. This thought is reinforced by the evidence that the Gisu are also well known for a kind of individualism that downplays social interdependence. Suzette Heald (1989), a later ethnographer of the Gisu, for instance, notes when commenting on the ritual of male circumcision which is a central feature of Gisu life:

> Though [circumcision] qualifies a boy for manhood, it is manhood that he does not really attain until he has established his own household independent of his father […] through his own efforts. […] Here as elsewhere in Gisu life, the onus is upon the individual. Gisu individualism […] systematically underplays the dependence of people on each other (p.77).

Victor Turner (1969) makes a similar observation when comparing the Gisu rites with those of the Ndembu: '..the Gisu rites are performed mainly for the individuals ….., while the Ndembu rites are collective affairs….. (p.235)'.

Individual Gisu belong to clans and sub-clans each with clear territorial boundaries. By contrast, individuals are also members of lineages who can live anywhere and who can enumerate their patrilineal links to their common ancestor, the legendary Masaba. Only the lineage is exogamous, is a source of kinship and entails precisely defined rights and obligations: the most important ones concern the negotiation and collection of bride wealth and the allocation of land rights when a boy is circumcised. Thus, like many people, the Gisu are connected to each other in more than one way (e.g. there are claims of territoriality as well as lineage kinship and there are complicated gender

relations). However, it also appears that these ties are rather weak. This is perhaps not so surprising given the earlier description of the Gisu as embodying a kind of 'unruly individualism'. Heal (1989) summarises:

> Not only are the most serious conflicts in Gisu life generated within the family and lineage but membership of such groups provides no countervailing set of loyalties which operate to give protection to the individual (p. 140).
>
> The Gisu man stands firmly alone, unable to commit his descent group either morally or in action on his behalf. […] A local man was well-known for his saying "there is not much in kinship" […] We are dealing here with a system of self-help that is so individualised that it differentiates little between kin and non-kin (p. 173).

Again, this aspect of Gisu life makes them well suited to our test of the experimental method because it would be surprising in these circumstances if they revealed the kind of ingroup favoritism that is usually observed in laboratory experiments.

The strong individualism and the relative weakness of kinship ties that are apparent from the ethnographic record were also revealed in the focus group discussions on trust in kin versus non-kin that we organized at the outset of the research; low levels of trust in kin were confirmed by participants and attributed to two major factors.[8] First, when land is divided among sons upon the death of a father, or when a boy is circumcised, signifying his entry into adulthood, the lineage – uncles and aunts – get involved. The proposed division is often contested and gives rise to conflicts, between the boy and his relatives, and, over time, among their descendants, as memories of perceived past injustices tend to persist. Heald (1989, 1999) finds a similar set of factors at play and explains their prominence by the population pressure: Bugisu is one of the most crowded places of rural Africa. Second, when families meet with misfortune, be it in the form of an untimely death, disease, infertility, harvest failure, or children failing at school, suspicions arise of better-off relatives having used witchcraft on them. These better-off relatives, in turn, fear retaliatory witchcraft.

II. Experiment

We use the standard Joyce Berg, John W. Dickaut, and Kevin A. McCabe (1995) trust game. One person (the *truster*) has an endowment of 5000 UGS and can choose to give some of it (=X) to the person they have been anonymously paired with. This gift is multiplied by 3 (=3X) and the recipient (the *trustee*) must decide how much, if any (=Y), to return to the original donor. The proportion of the endowment that is given by the truster is the *giving rate* and is taken as the index of trust (i.e. = X/5000). The proportion that is returned by the trustee of what he or she receives is the *return rate* and is treated as a measure of his or her trustworthiness (i.e. = Y/3X). The trust game was played under 4 different treatments with males (to avoid the complicating factor of gender relations).

> Treatment A: each person knows that they are paired with a Gisu man who lives in the Manafwa district.
>
> Treatment B: each person knows they are paired with a Gisu man who lives in their own village.

[8] We organised four focus group discussions among members of the same lineage only, and six focus group discussions with participants that each belonged to a different lineage; a focus group consisted of between 10 and 14 participants.

Treatment C: each person knows that they are paired with a Gisu man who belongs to their own village and their own lineage.

Treatment D: each person knows that they are paired with a Gisu man who lives in a different village and who belongs to a different lineage to their own.

The move from A to B strengthens the ties of territoriality between the players and this is reinforced in C by the addition of common lineage. In this way, the contrast between behavior in A, B and C tests for how behavior is affected by strengthening the 'own group' ties between players. Treatment D exactly reverses the conditions found in C: each player now knows they interacting with someone who comes from another village and belongs to a different lineage to their own. As a result, the contrast between behavior in C and D reveals the difference in treatment between fellow members of a group and those between members of different groups.

We recruited subjects from Namabya, a sub-county, of the Manafwa District[9] in Bugisu. They have a population 18,000 inhabitants[10] and was selected because, unlike many other areas in the region, it has remained predominantly Gisu.[11] 10 sufficiently large villages were chosen with the sole aim of realizing a maximum geographical spread to avoid the possibility of between-day contamination. The population size of these villages ranged from 400 to 600 individuals. In each village, a census was taken of all Gisu males over 18 years old and circumcised – their lineage was indicated as well as the household they belong to. We selected only male subjects, to aid with interpretation in the presence of strong gender differentiations. Among the lineages with a sufficient number of eligible males, one was randomly selected, and fourteen (two spare) eligible males were randomly selected subject to the restriction of at most one per household. Among all other eligible males, fourteen (again two spare) were randomly selected, subject to the same restriction of at most one per household. In the weeks preceding the experiment, the selected individuals were personally invited. If they were unavailable (very rare) the person adjacent on the list (belonging to the same lineage where appropriate) was invited. If all fourteen lineage members/others showed up, the two last-comers were paid the show-up fee and sent home. In no case, did fewer than twelve show up.

The game was piloted among 30 subjects. The script was read out slowly and deliberately and participants were instructed individually and were given ample opportunity to ask questions of clarification. Understanding of the instructions appeared to be satisfactory in that most players in the actual experiments, 89.58 percent, answered the post-game control question correctly.[12] In order to avoid players in treatments A and D recognising the village and/or lineage of their co-players, it was decided to visit players in their village and to communicate offers and counter-offers by mobile phone.[13] On each of five game days, two villages were matched: subjects were unaware that they were matched with only one other village, unless their specific experimental instructions made that explicit (in treatment D). Every day, 24 villagers played per village (the typical venue was a village

[9] Manafwa District is located in the Eastern Region of Uganda. It borders on the Republic of Kenya to the east, the district of Bududa in the North, Mbale town to the west and Tororo town to the south-west. It has a land area of about 451 sq km, and a population of about 315,000 (source: Government of Uganda, 2003). The district is largely agricultural. The average rainfall is 1,500mm per annum, and supports intensive agriculture. Main food crops grown in the district are bananas, maize and beans; main cash crops are coffee and cotton. (Source: key informants).

[10] Source: Government of Uganda (2003).

[11] Due to its remoteness, some 50 km on dirt roads from Mbale town, the nearest town, there is relatively little immigration.

[12] The following post-game control question was put to Players 1, "If you had decided to send 4,000 shillings to the other person, which we would then have tripled; and the other person sent back 2,000 shillings – what would have been your earnings?" and to Players 2, "If the other person had decided to send 4,000 shillings, which we would then have tripled; and you had sent back 2,000 shillings – what would have been your earnings?"

[13] In some cases the network let us down and offers were communicated by research assistants walking to a pre-arranged meeting point – this eventuality was anticipated, and taken into account in deciding which villages were to be matched – these were never more than 2 km. apart.

hall), so 48 in total per day, and 240 (60 per treatment) in total. Per village, six of these played treatment A, six treatment B, six C and six D – within each treatment, three were randomly assigned the role of Player 1, three that of Player 2.

III. Results

Figure 1 sets out the broad results by treatment. The average giving rate is lowest in treatment A (25%) and highest in treatment D (32.1%), with treatments B and C somewhere in the middle (27.6% and 28.5%, respectively). The average return rate is 25.5%, 27.6%, 25.3% and 35.5% in treatments A, B, C and D respectively.

(*Insert Figures 1 and 2 about here*.)

Figure 2 compares these giving and return rates with those found in the 28 trust studies reviewed in Juan Camilo Cardenas and Jeffrey Carpenter's (2008) metastudy of trust game experiments in developing countries. The average return rate in those studies was 38.1%, and in 15 of those studies return rates are above 40%. The difference is also statistically significant (Mann Whitney P = 0.04, two tailed), although in the case of treatment D the value is close to the average. The results are more striking for average giving rates: Figure 2 shows that only 2 of 28 of the other studies found giving rates comparable to those observed in our experiment (Mann Whitney P = 0.002, two tailed).

RESULT 1: *There is significantly less trust and trustworthiness among the Gisu than is normally observed.*

We now focus on the across treatments differences. The difference in giving rate between treatments A and D is statistically significant (Mann-Whitney P = 0.048, two tailed), and is in the direction of outgroup rather than ingroup favoritism: agents gave more when they knew they were paired with someone from a different village and lineage. While the difference in return rates between treatment A and D is not statistically significant in a two tailed test (Mann Whitney P = 0.103), and there are not other statistically significant coefficients, it is apparent that once again the highest point return rate is in treatment D where agents knew they were paired with someone from a different village and lineage. Again, there is clearly no evidence of ingroup favoritism.

It is possible that the lack of apparent ingroup favoritism is due to the range of other covariates that may differ across subjects. For example, in considering return rates, we need to take into account that agents may behave reciprocally, thus returning proportionately less to subjects who give them less (as observed, for example, in Hargreaves Heap and Zizzo, 2009), and, given the pattern of giving across treatments, this may obscure any tendency that agents to exhibit ingroup favoritism in their return decisions. To control for this and other such problems from elementary bivariate Mann Whitney tests, we ran a number of regressions on both giving rates and return rates.

(*Insert Table 1 about here*.)

Table 1 reproduces the results of OLS regressions controlling for error clustering at the village level (Model 1 for the giving rate, Model 1' for the return rate) and of robust regressions controlling for outliers and variously handling village level specific effects (Models 2-4 for the giving rate, Models 2'-4' for the return rate).[14] The independent variables in the giving rate regressions are the

[14] We also ran a number of other specifications, including random effects regressions, tobit random effects regressions and quantile regressions; we also considered reduced sets of variables. The key qualitative results as described in this section remain robust: e.g., there were never negative point coefficients on the treatments D coefficients in either giving rates or return rates, and more generally there was no evidence of ingroup favoritism.

treatment dummies (B, C and D, = 1 for subjects in treatments B, C and D respectively), age, NMales (the number of males in the household aged between 16 and 60), NFemales (the number of females in the household aged between 16 and 60), Land (lanholding in acres: a measure of wealth), NOlderSiblings (the number of older siblings), NYoungerSiblings (the number of younger siblings), ControlQ (a dummy equal to 1 if the subject correctly responded to a control questionnaire to check for understanding) and Illiterate (a dummy equal to 1 if the subject cannot read). The return rate regressions also include GivingRate (the giving rate the subject has benefited from) and interaction terms between the GivingRate and the treatment dummies (B × GivingRate, C × GivingRate and D × GivingRate).

In the giving rate regressions, the only weak evidence of potential ingroup favoritism is in relation to treatment B relative to treatment A ($\beta = 0.094$, $P = 0.068$). However, once we control for village level effects (Models 1, 3 and 4), even this marginal significance disappears. Furthermore, the treatment D coefficient is robustly positive across all giving rate regressions (suggesting outgroup favoritism), albeit statistically insignificantly so in Model 1 ($P = 0.27$), marginally significantly so in Model 3 ($P = 0.074$) and more clearly so in Models 2 ($P = 0.027$) and 4 ($P = 0.05$).

RESULT 2: *There is almost no evidence of ingroup favoritism in trust; if anything, the evidence points to greater giving to people belonging to different lineages and villages.*

In the return rate regressions, the treatment dummies are statistically insignificant, though the treatment D point coefficient remains always positive and higher than those on treatments B and D. Interestingly, we also find no evidence of reciprocation, with only exception of treatment C, where, once village level effects are controlled for has a statistically significant positive coefficient ($P = 0.017, 0.059, 0.028$ in models 1', 3', 4'). So there *is* some evidence of reciprocity if subjects know they are interacting with someone from the same lineage and village, as in treatment C, but, as shown by Figure 1, this does not lead to greater average trustworthiness.

RESULT 3: *There is no evidence of ingroup favoritism in trustworthiness.*

RESULT 4: *We find no general evidence of reciprocity, with the only potential exception of when agents know that they are interacting with people belonging to the same lineage and village (although, even then, average trustworthiness does not change).*

IV. Discussion and Conclusion

If the laboratory evidence of pro-sociality and ingroup favoritism is in some important respects an artifact of the experimental conditions themselves rather than an indication of how people behave outside the laboratory, then one more experiment that reveals pro-sociality and ingroup behaviors on a population, known in some degree to exhibit these properties outside the laboratory, hardly tells in favor of external validity. This is why severe tests of theories are appealing. The serious attempt at falsifying the claim of external validity in these circumstances comes instead from an experiment on a population that is known not to exhibit readily these behaviors. The Gisu appear to be such a population. And our hypotheses, based on a belief in the external validity of the experimental method, are that an experiment with the Gisu will i) reveal less pro-social behavior than is usual in experiments and ii) where pro-sociality does arises, insiders will not be favored.

Neither hypothesis is rejected and, on this account of the matter, the experiment supplies a serious, but unsuccessful, attempt at falsifying the belief in the external validity of experiments; and this is as good as it gets when 'falsificationism' rather than 'verificationism' drives science. This is not to suggest, of course, that 'verification' should play no role in the assessment of external

validity (or that the careful elaboration of experimental conditions cannot be used to examine how behavior is likely to change outside the lab). Rather our claim is that, in addition to these approaches, severe tests of theories should form part of the strategy for assessing the external validity of experiments; and we provide an example.

REFERENCES

Andreoni, James, William T. Harburgh and Lise Vesterlund. 2007. "Altruism in Experiments." The New Palgrave Dictionary of Economics, 2nd Edition.

Assudani, Rashmi H. 2009. "Ethnic Entrepreneurship: The Distinct Role of Ties." *Journal of Small Business and Entrepreneurship*, 22(2): 197-205.

Berg, Joyce, John W. Dickaut, and Kevin A. McCabe. 1995. "Trust, Reciprocity and Social History." *Games and Economic Behavior*, 10(1), 122-42.

Cardenas, Juan Camilo and Jeffrey Carpenter. 2008. "Behavioral Development Economics: Lessons from Field and Labs in the Developing World." *Journal of Development Studies*, 44(3), 311-338

Charness, Gary, Luca Rigotti, and Aldo Rustichini. 2007. "Individual Behavior and Group Membership." American Economic Review, September, 97(4), 1340-1352.

Guala, Francesco. 2005. *The Methodology of Experimental Economics*. Cambridge: Cambridge University Press.

Hargreaves Heap, Shaun and Daniel J. Zizzo. 2009. "The Value of Groups." *American Economic Review*, 99(1), 295-323.

Heald, Suzette. 1989. *Controlling Anger: The Anthropology of Gisu Violence*, Oxford: James Currey.

Heald, Suzette. 1999. *Manhood and Morality – Sex, Violence and Ritual in Gisu Society*, London and New York: Routledge.

Kuhn, Thomas S. 1970. *The Structure of Scientific Revolutions*, 2nd Ed. Chicago: University of Chicago Press.

La Fontaine, Jean. (1959). "The Gisu of Uganda" in: Daryll Forde (ed.), Ethnographic Survey of Africa, Part X, East Central Africa, London: International African Institute.

Levitt, Steven D. and John A. List. 2007. "What Do Laboratory Experiments Measuring Social Preferences Reveal About the Real World?" *Journal of Economic Perspectives*, 21(2), 153-74.

Mayer, Adalbert, and Steven L. Puller. 2008. "The Old Boy (and Girl) Network: Social Network Formation on University Campuses." *Journal of Public Economics* 92(1-2): 329-347.

Mayo, Deborah G. 1996. *Error and the Growth of Experimental Knowledge*. Chicago: University of Chicago Press.

Mayo, Deborah G., and Aris Spanos. 2006. "Severe Testing as a Basic Concept in a Neyman-Pearson Philosophy of Induction." *British Journal of the Philosophy of Science*, 57(2): 323-357.

Popper, Karl. 1963. *Conjectures and Refutations*. London: Routledge and Kegan Paul.

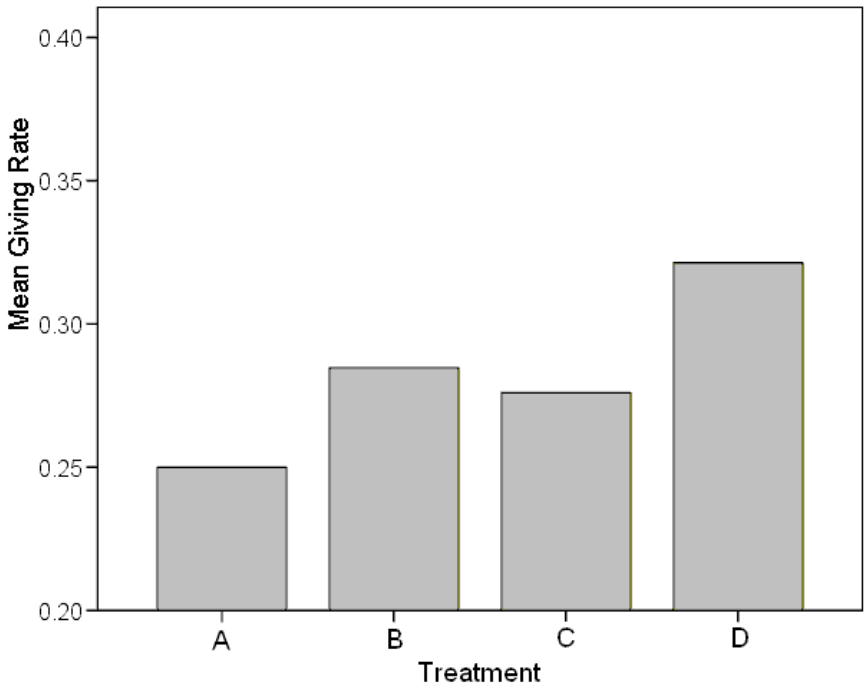
Putnam, Robert D., Robert Leonardi and Rafaella Y. Nanetti. 1993. Making Democracy Work Civic Traditions in Modern Italy. Princeton and Chichester: Princeton University Press.

Turner, Victor. 1969. "Symbolization and Patterning in the Circumcision Rites of Two Bantu-speaking Societies" in: Mary Douglas and Phyllis M. Kaberry (ed.), Man in Africa, London: Tavistock Publications.

.

Figure 1. Mean Giving and Return Rates by Treatment

(a) Mean Giving Rates



(b) Mean Return Rates

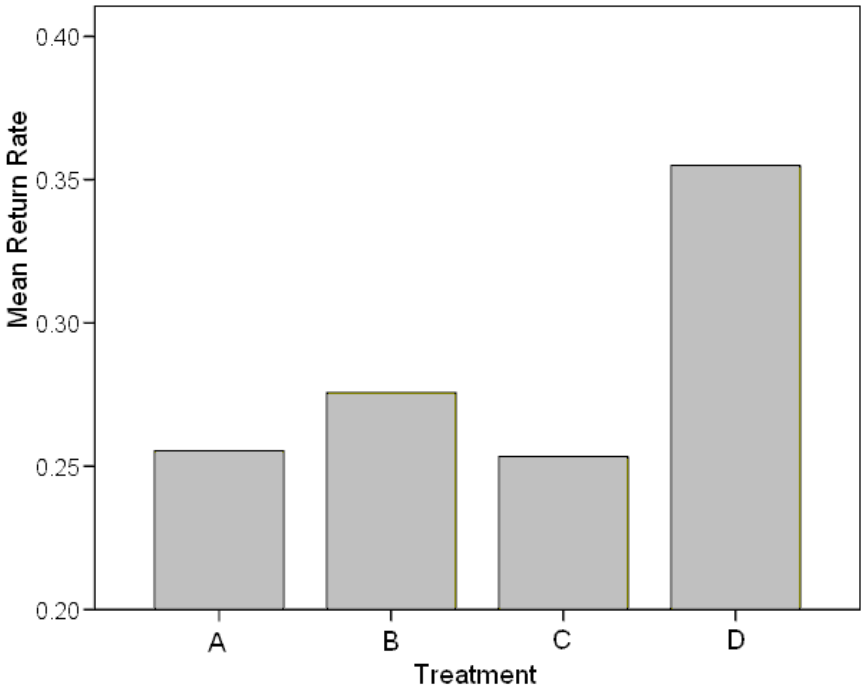
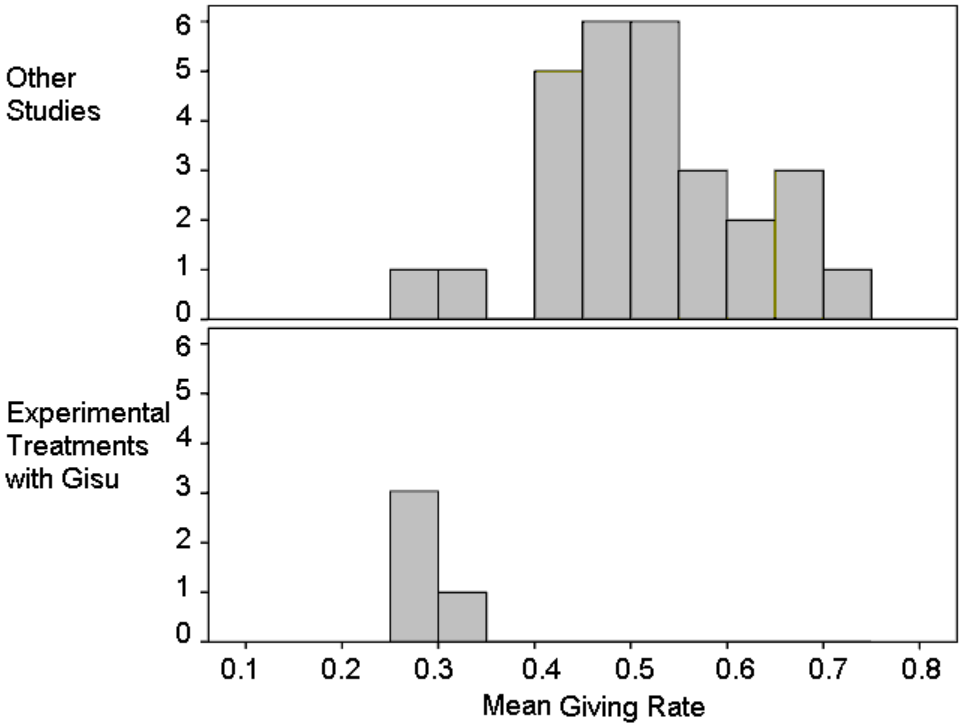
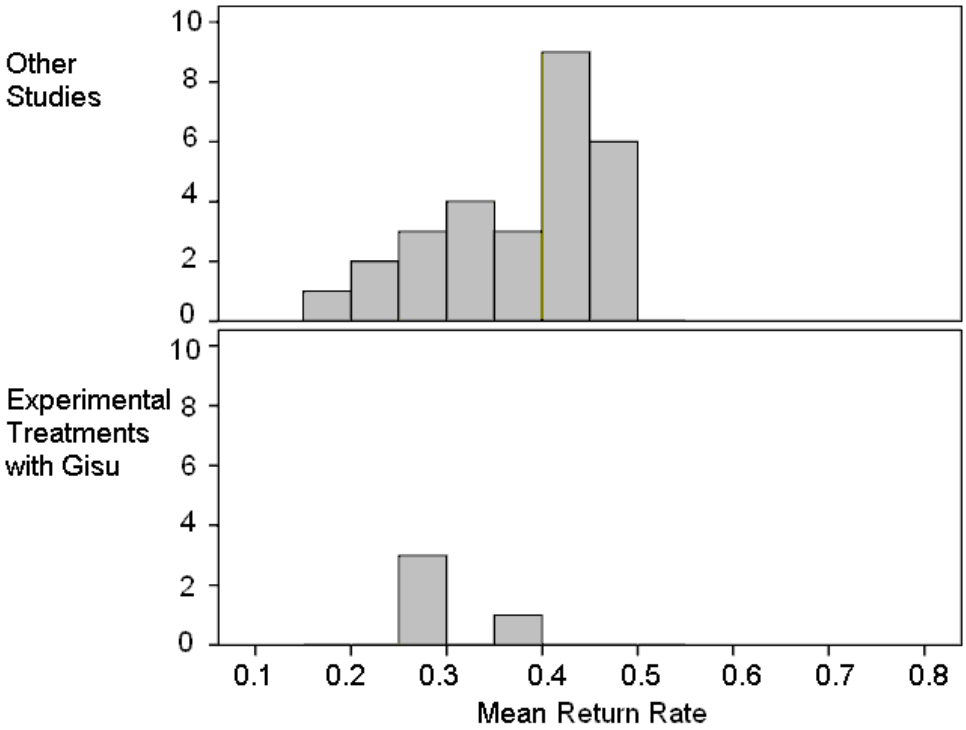


Figure 2. Histograms of Mean Giving and Return Rates across Studies

(a) Mean Giving Rates



(b) Mean Return Rates



Notes: Data with reference to other studies has been taken from Table 1 of Cardenas and Carpenter (2008, pp. 314-315; $n = 28$).

Table 1. Regressions on Giving Rates and Return Rates

	Regressions on Giving Rate											
	Model 1			Model 2			Model 3			Model 4		
	β	*t*	*P*	β	*t*	*P*	β	*t*	*P*	β	*t*	*P*
Treatment B	0.054	0.8	0.445	0.094	1.84	0.068	0.074	1.43	0.155	0.074	1.49	0.14
Treatment C	0.024	0.29	0.778	0.046	0.9	0.372	0.038	0.73	0.47	0.043	0.85	0.397
Treatment D	0.078	1.18	0.27	0.118	2.24	0.027	0.096	1.8	0.074	0.101	1.98	0.05
Age	0.001	1.42	0.19	0.001	0.93	0.354	0.002	1.22	0.226	0.001	0.92	0.359
NMales	-0.009	-0.55	0.596	-0.003	-0.2	0.839	-0.005	-0.35	0.73	-0.005	-0.33	0.74
NFemales	0.032	2.03	0.073	0.028	1.99	0.049	0.019	1.25	0.213	0.016	1.16	0.25
Land	-0.009	-0.63	0.543	0.003	0.22	0.83	0	0.03	0.975	0.001	0.07	0.94
NOlderSibilings	0.004	0.91	0.385	0.009	1.19	0.237	0.003	0.32	0.751	0.004	0.56	0.578
NYoungerSiblings	0.006	0.51	0.626	0.006	0.68	0.499	0.004	0.45	0.652	0.002	0.25	0.805
ControlQ	-0.017	-0.29	0.78	-0.04	-0.69	0.491	-0.069	-1.03	0.307	-0.029	-0.51	0.61
Illiterate	-0.044	-1.3	0.227	-0.018	-0.47	0.638	-0.016	-0.4	0.694	-0.029	-0.79	0.432
Constant	0.182	1.9	0.09	0.113	1.29	0.199	0.078	0.73	0.468	0.133	1.52	0.132
Village Dummies	No			No			Yes			Selected		
	Regressions on Return Rate											
	Model 1'			Model 2'			Model 3'			Model 4'		
	β	*t*	*P*	β	*t*	*P*	β	*t*	*P*	β	*t*	*P*
Treatment B	0.08	0.93	0.374	0.106	0.91	0.364	0.075	0.67	0.504	0.063	0.6	0.553
Treatment C	-0.123	-1.33	0.216	-0.099	-0.91	0.367	-0.17	-1.64	0.106	-0.189	-1.88	0.064
Treatment D	0.169	1.28	0.233	0.11	0.88	0.383	0.153	1.29	0.2	0.106	0.92	0.362
GivingRate	0.021	0.22	0.829	0.002	0.89	0.378	0.067	0.33	0.746	0.007	0.04	0.971
B x GivingRate	-0.131	-0.72	0.491	0.06	0.29	0.772	-0.23	-0.72	0.473	-0.153	-0.52	0.604
C x GivingRate	0.441	2.91	0.017	-0.228	-0.71	0.479	0.585	1.92	0.059	0.639	2.24	0.028
D x GivingRate	-0.211	-1.06	0.318	0.451	1.45	0.15	-0.306	-1	0.319	-0.158	-0.54	0.591
Age	0	-0.24	0.818	-0.139	-0.43	0.665	0.001	0.43	0.667	0.001	0.43	0.672
NMales	0.003	0.1	0.926	-0.018	-1.09	0.279	-0.007	-0.42	0.677	-0.008	-0.54	0.592
NFemales	0.035	1.19	0.264	0.038	2.1	0.038	0.013	0.7	0.487	0.02	1.16	0.251
Land	-0.028	-1.47	0.177	-0.028	-2	0.049	-0.019	-1.36	0.179	-0.021	-1.59	0.116
NOlderSibilings	0.001	0.11	0.915	0	0.05	0.962	0.005	0.54	0.589	0.002	0.21	0.833
NYoungerSiblings	0	0.02	0.985	0.002	0.17	0.863	-0.011	-0.95	0.345	-0.011	-0.95	0.344
ControlQ	-0.132	-2.6	0.029	-0.171	-2.41	0.018	-0.1	-1.46	0.149	-0.123	-1.87	0.065
Illiterate	0.027	0.54	0.6	0.056	1.23	0.221	0.018	0.45	0.655	0.019	0.45	0.655
Constant	0.349	2.8	0.021	0.261	2.07	0.042	0.294	2.39	0.019	0.32	2.76	0.007
Village Dummies	No			No			Yes			Selected		

Notes: $n = 118$ for giving rate regressions, $n = 100$ for return rate regressions.[15] Models 1 and 1' are OLS regressions allowing for village level error clustering. Models 2-4 and 2'-4' are robust regressions controlling for outlier observations; models 3 and 3' include village level dummies to also control for village level effects; model 4 (4') includes only the village dummies that display $P < 0.2$ in regression 3 (3', respectively).

[15] $n < 120$ because of occasional missing variables; n is lower for return rate regressions because observations are missing if a trustee has been given 0 and therefore is unable to return.