Economics Letters 114 (2012) 308–311

Contents lists available at SciVerse ScienceDirect

Economics Letters

journal homepage: www.elsevier.com/locate/ecolet

Exploring whether behavior in context-free experiments is predictive of behavior in the field: Evidence from lab and field experiments in rural Sierra Leone

Maarten Voors [a,c], Ty Turley [b], Andreas Kontoleon [c], Erwin Bulte [a,d,*], John A. List [b]

[a] Development Economics Group, Wageningen University, P.O. Box 8130, 6700 EW Wageningen, The Netherlands
[b] Department of Economics, The University of Chicago, United States
[c] Department of Land Economy, University of Cambridge, United Kingdom
[d] Department of Economics, Tilburg University, The Netherlands

ARTICLE INFO

Article history:
Received 28 July 2011
Accepted 14 October 2011
Available online 21 October 2011

JEL classification:
C9
D03
O12
O55

Keywords:
Lab games
Field experiments
Community driven development
External validity
Pro-social preferences
Africa

ABSTRACT

We use a sample of subsistence farmers in Sierra Leone as respondents to compare behavior in a context-free experiment (a standard public goods game) and behavior in the field (a real development intervention). There is no meaningful correlation in behavior across contexts. This casts doubt on the prospect of using lab experiments as "predictors" of behavior in real life.

1. Introduction

Important questions have been raised about the external validity of laboratory-style experiments. In the context of altruism and public good giving, for example, Laury and Taylor (2008) ask whether decisions made in sterile lab settings using tokens, anonymous partners, and benefits accruing to a small number of players (those concurrently in the experiment) are predictive of altruistic behavior in naturally occurring settings. They conclude "one should be cautious when using the results from laboratory public good experiments to make inferences about altruism outside of the laboratory" and argue that "further investigation into the external validity of decisions made in context-free situations is warranted" (p. 28). We agree. If behaviors across context are correlated, behavior in the lab would be predictive of behavior in real life. We would then be able to credibly "testbed" real policies and interventions in the lab, at modest cost, and could identify individuals for targeting by such interventions (or weed out free riders).

The evidence with respect to the external validity of lab-style experiments is mixed. Some studies document positive correlations between behavior in the game and real life (Karlan, 2005; Benz and Meier, 2008; Bouma et al., 2008; Carpenter and Seki, 2010) while others reject consistency in behavior across contexts (e.g. List, 2006; Voors et al., 2011). In what follows, we focus on one specific form of behavior—contributions to a (local) public good. Such contributions are usually believed to reflect pro-social preferences, most notably altruism. Laury and Taylor (2008) find that pro-social behavior in the lab predicts contributions to a local public good, but not in a uniform way. Specifically, while some measures of altruism based on standard public good (PG) games in the lab are correlated with the likelihood of acting altruistically in real life, the same is not true for other measures.

Why may respondents' behavior differ across contexts? Levitt and List (2007) propose that such divergence may be explained by several factors, including the presence or absence of moral and ethical considerations, the subject pool, the context in which the choice is embedded, the extent to which one's actions are scrutinized by others and the nature of that scrutiny; and the stakes of the game. These factors may serve as shift parameters. Obviously, a subject's contribution relative to others in her group (her "ranking") across lab and field settings are preserved if these shifters are isomorphic across people (even if the "levels" of

* Corresponding author. Tel.: +31 317 484879; fax: +31 317 484037.
E-mail addresses: maarten.voors@wur.nl (M. Voors), tyturley@gmail.com (T. Turley), ak219@cam.ac.uk (A. Kontoleon), erwin.bulte@wur.nl (E. Bulte), jlist@uchicago.edu (J.A. List).

0165-1765/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.econlet.2011.10.016

Table 1
Explaining real life altruism by altruism displayed in the lab.

	(1) PG Aid	(2) PG Aid	(3) PG Aid	(4) Community Labor	(5) Community Labor	(6) Community Farm	(7) Community Farm
PGG	−161.8	439.8	−1701.9	−0.230	0.396	−0.0710	−2.075
	(943.3)	(2013.1)	(9520.7)	(0.383)	(1.466)	(0.419)	(1.431)
Household controls	No	Yes	Yes	Yes	Yes	Yes	Yes
Interaction terms	No	No	Yes	No	Yes	No	Yes
Village fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	453	111	111	115	115	113	113
adj. R^2/Pseudo R^2	0.13	0.17	0.19	0.19	0.19	0.18	0.21

Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Household controls are age, education, household income, conflict exposure in the civil war, and witchcraft beliefs. The full set of interaction terms are the product of these household controls and the PGG variable. Columns (1)–(3) are estimated using an interval model and Columns (4)–(7) are estimated using ordered logit.

pro-social preferences displayed will vary). In contrast, allowing heterogeneity in shift parameters implies that we may observe different relative rankings across lab and field.

We revisit the question whether behavior in the lab spills over to the field. Our analysis differs from earlier studies because of at least three factors. First, we use an under-researched yet highly relevant pool of respondents—poor villagers from small and remote forest edge communities in Sierra Leone, increasingly subjected to community-based interventions by NGOs, often explicitly engaged in attempts at social engineering. Second, we compare choice behavior in related dilemma situations by inviting subjects to participate in a lab-style public goods game and a development intervention with similar structure (in our case: involving the allocation of an endowment between self and others). Important differences between the lab game and field experiment are that in the latter (i) the stakes are much higher (a month of wages versus a day's wage), and (ii) the public good accrues to a larger number of people (affecting the entire village instead of two other co-players). Third, we also compare experimental play with survey-based evidence on actual public good contributions.

2. Experiments and data

In the summer of 2010 we visited 500 randomly sampled households in 25 villages in rural Sierra Leone. These respondents participated in a livelihoods intervention implemented by an international NGO, and a standard public goods game. In the livelihood intervention, respondents were endowed with $20 (or 80,000 Leones, Le), and asked to divide this endowment between private goods for themselves and a community project fund for the entire village. On average, participants allocated 75% of their endowment to private goods and 25% to the community project fund.

After completing the intervention, we played a conventional five round PG game where respondents were randomly and anonymously matched with two peers.[1] Players were endowed with five tokens. Tokens kept were worth 1000 Le, and tokens "invested" were worth 500 Le to all group members. The payoff under full co-operation thus amounts to 7500 Le, and the maximum possible payoff for a free rider (assuming full contributions by his peers) was 10,000 Le. On average, households invested two tokens (40%) in each round, and the average payoff was 6000 Le.

Next, we collected additional data using household and village level surveys in the same 25 villages (but now involving only a subsample of 170 respondents). We collected demographic, socio-economic, institutional and civil war information as control variables, and also used the household survey to construct additional dependent variables. As survey-based proxies for pro-social preferences, we use the degree to which the respondent contributes labor to community projects or to the village farm (both measured on a 4-point scale).[2]

Our main identification strategy is simple: we use an interval regression model and regress our measures of "real life" pro-social preferences Z_{ij}—contributions in the livelihood intervention by individual i ($i = 1, \ldots, 500$) in village j ($j = 1, \ldots, 25$) as well as measures of stated behavior—on lab data (PGG_{ij}), household controls (X_{ij}) and village fixed effects (V_j)[3]:

$$Z_{ij} = \alpha + \gamma_1 PGG_{ij} + \gamma_2 X_{ij} + \gamma_3 V_j + \mu_{ij}^{FE}, \quad (1)$$

where μ is an error term. To explore whether there is evidence of systematic heterogeneity in shift parameters, we also estimated models with a full set of interaction terms:

$$Z_{ij} = \alpha + \gamma_1 PPG_{ij} + \gamma_2 X_{ij} + \gamma_1 X_{ij} PPG_{ij} + \gamma_3 V_j + \mu_{ij}^{FE}. \quad (2)$$

Finally, we have undertaken a series of non-parametric tests. Based on their choices in the PG game we divide our sample of respondents in 3 subsamples, distinguishing between (weak) free riders (respondents contributing 1 token or less in all five rounds)[4], mild cooperators (consistently contributing 2 or 3 tokens) and altruists (always contributing 4 or 5 tokens). We then explore whether the behavior across these groups in the livelihoods intervention was different, or not.

3. Results

Table 1 summarizes the regression results. The main result is that behavior in the lab game is not correlated with our measures of pro-social preferences in the field. In some models, the correlation is even negative, but standard errors are so large that lab play never enters significantly. This is true when explaining choices in the aid intervention experiment (columns 1–3) as well as the

[1] In each village the aid experiment preceded the public goods game. Subsequently, ordering effects may bias the level of contributions though it is not evident that ordering affects the ranking of individuals. In a similar experiment Laury and Taylor (2008) document that experimental order does not affect their results.

[2] Local public goods provision is heavily dependent on the private provision of labor by community members. Community Farm is a variable measured on a 1–4 scale, where a higher score means subjects contribute more labor to the common farm, the revenues of which accrue to the community as a whole for festivities or visiting dignitaries. Community Labor was also measured on a 1–4 scale with a higher score implying that subjects contribute more labor to community projects, such as road clearing, water flow maintenance, village clean up, and other community projects.

[3] We are interested in the (conditional) correlation between our pro-sociality measures, and the regression structure is *not* intended to suggest that we are looking for a causal effect of lab behavior on behavior in the field. To emphasize this point: we have also estimated a linear seemingly unrelated regression (SURE) model to test whether the observed responses in the experiments are correlated (see below).

[4] Note that we classify our subjects based on behavior throughout the five PG game rounds. Dropping the last round (when subjects have a reduced incentive to make pro-social choices) does not change the results.

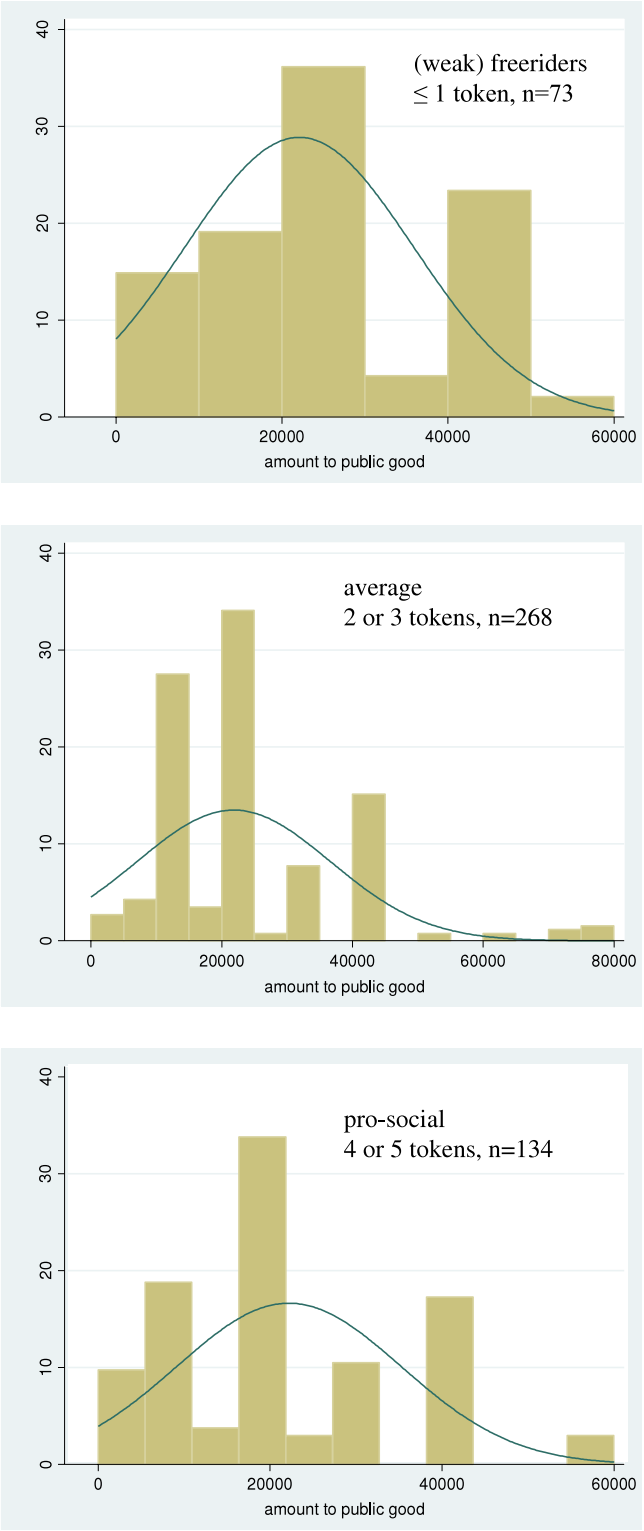

Fig. 1. Non-parametric analysis of altruism in the lab and in the field.

voluntary contributions to community projects (columns 4–7). It does not matter whether we control for household or respondent characteristics (age, education, income, etc.), or not. Also, including a full set of interaction terms – household controls times behavior in the lab game – does not affect this result. Interestingly, the interaction terms tend to enter insignificantly, suggesting we cannot capture heterogeneity across respondents in such a simple fashion (results not shown, but available on request).

The lack of correlation between behaviors is robust. Our linear SURE model, used to test whether the observed responses in the experiments are correlated, yields similar results: the error terms of the models explaining behavior in the lab game and in the framed field experiment are not correlated (details available on request). This lack of correlation is also confirmed in the non-parametric analysis. Fig. 1 reports behavior in the livelihoods intervention for our three sub-groups of respondents (weak free riders, mild cooperators, and altruistic respondents). The distribution of contributions in the intervention does not vary across groups—this is confirmed by formal tests.[5] Taking the evidence together we conclude that, for our pool of respondents, play in lab experiments has no predictive power for behavior in naturally occurring settings. Indeed, the SURE model suggests the lab game may measure a distinct latent variable altogether.

4. Discussion

Why are pro-social preferences measured in the lab not correlated with pro-social preferences measured in real life, where salient choices are made in a familiar context? Our analysis does not allow us to identify the non-isomorphic shifter that is responsible for this finding—doing so would require a more elaborate experimental design in which we randomly vary different candidate shifters (such as the nature of scrutiny, or the stakes of the choice experiment) while keeping other shifters constant. This is left for future work.

Nevertheless, we believe the results summarized on these pages are of interest. They are based on a particular yet relevant subject pool (adding to the generality of the main insight, also obtained in "more conventional" settings). The lack of correlation suggests aid agencies considering lab games as a cheap way to help them design their policies should think twice. The robustness of our main result is strengthened by the fact that the game and real behavior settings are so similar, and the fact that we obtain similar results using survey-based evidence on actual public good contributions.

Our results are consistent with the finding in psychology that behavior in different settings is only weakly correlated (even across lab settings; see, e.g., Ross and Nisbett (1991) for a review). This does not mean that there is necessarily something wrong with one of the experimental mechanisms in this study. Rather, it may mean they are measuring different latent variables or preferences—perhaps a general cross-situational preference for pro-social behavior does not exist?

Acknowledgments

We are indebted to the UK's Royal Society for the Protection of Birds (RSPB), the Gola Forest Programme (supported by the European Union, Fonds Française pour l'Environnement Mondial and the Global Conservation Fund at Conservation International), BirdLife International, Prof. Paul Richards, Esther Mokuwa and Koen Leuveld for their collaboration in this project. We thank N.W.O. 452-04-333 and Cambridge Conservation Initiative (CCI 05/101005) for financial support.

[5] Two-sided between group comparisons are all non-significant. There is a residual category of respondent who make inconsistent choices through the five rounds of the experiment, their contributions in the aid experiment are also not significantly different from the contributions of subjects in the other groups.

References

Benz, M., Meier, S., 2008. Do people behave in experiments as in the field? evidence from donations. Experimental Economics 11, 268–281.

Bouma, J., Bulte, E., van Soest, D., 2008. Trust, trustworthiness and cooperation: social capital and community resource management. Journal of Environmental Economics and Management 56, 155–166.

Carpenter, J., Seki, E., 2010. Do social preferences raise productivity? Field experimental evidence from fishermen in Toyama Bay. Economic Inquiry 49, 612–630.

Karlan, D., 2005. Using experimental economics to measure social capital and predict financial decisions. American Economic Review 95, 1688–1699.

Laury, S.K., Taylor, L.O., 2008. Altruism spillovers: are behaviors in context-free experiments predictive of altruism toward a naturally occurring public good? Journal of Economic Behavior and Organization 65, 9–29.

Levitt, S., List, J.A., 2007. What do laboratory experiments measuring social preferences reveal about the real world? Journal of Economic Perspectives 21, 153–174.

List, J.A., 2006. The behavioralist meets the market: measuring social preferences and reputation effects in actual transactions. Journal of Political Economy 114, 1–37.

Ross, Lee, Nisbett, Richard E., 1991. The Person and the Situation; Perspectives of Social Psychology. McGraw-Hill, New York.

Voors, M., Turley, T., Kontoleon, A., Bulte, E., List, J., 2011. Using artefactual field experiments to learn about the incentives for sustainable forest use in developing economies. American Economic Review 101, 329–333. Papers and proceedings.