CBESS Discussion Paper 10-03

Can Experimental Measures of Sensitivity to Social Pressure Predict Public Good Contribution?

by Daniel John Zizzo* and Piers Fleming

*University of East Anglia



Public good contribution in experiments may at least partially be driven by the social demand to contribute that is implicit in them. We consider a questionnaire measure and build a behavioural measure of sensitivity to social pressure based on paired dictator and money burning games; we find that the two are related. The evidence for social demand effects on public good contribution is mixed.

JEL classification codes

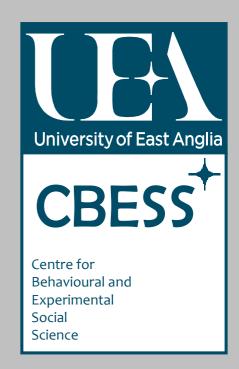
C91, D03, D12, D21, L11

Keywords

Social pressure, experimenter demand effects, public good contribution, dictator games, money burning

Acknowledgement:

Thanks to the Nuffield Foundation for funding this research (SGS/35859) and to Kei Tsutsui and Fabio Galeotti for his research assistance. Thanks also for advice and encouragement to presentations at the Economic Science Association meeting, Innsbruck (Austria), the Subjective Probability, Utility and Decision Making conference, Rovereto (Italy), and at the University of East Anglia, where versions of this work were presented. Experimental instructions for this research can be found online at Daniel John Zizzo's homepage (http://www.uea.ac.uk/~ec601/).



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

Contact details:

Daniel John Zizzo, School of Economics and CBESS, University of East Anglia, Norwich, UK

Phone: +44-1603-593668; Fax: +44-1603-456259

d.zizzo@uea.ac.uk

Piers Fleming, School of Social Work and Psychology and CBESS, University of East Anglia, Norwich, UK *p.fleming@uea.ac.uk*

1. Introduction

Social pressure is an obvious potential source of contribution in public good settings (e.g., Amos, 1982). In the context of public good experiments, it may be connected to a general social norm to cooperate. If this is the case, then results from public good experiments, rather than being attributable to social preferences such as inequity aversion (Fehr and Schmidt, 1999) or reciprocity (Falk and Fischbacher, 2006), may be largely an artifact of the social demands of the task faced by economic agents. Our note describes an experiment testing this intuition. To put it differently, it tests what Zizzo (2009) labels social experimenter demand effects, though social pressure as a source of public good contribution is obviously of broader relevance and interest than simply what may be happening in the experimental laboratory. It also verifies the extent to which social pressure sensitivity measures are able to predict how purely social punishment may be effective in raising contribution in public good settings, as in Masclet et al. (2003).¹

Our questionnaire measure relies on the psychology concept of social desirability. Social desirability is the idea that some people's reports of their attitudes or behavior are skewed towards over-reporting positive characteristics (e.g. voting) and under-reporting negative characteristics (e.g. weight, Larson, 2000). Psychologists measure social desirability by the means of multi-item questionnaire scales which ask about a selection of social desirable but atypical behaviors e.g. 'I always stay friendly and courteous with other people even when I'm stressed out' (Stober, 2001). The Stober (2001) 17 items scale is the one used in our experiment.

Our behavioral measure is new but inspired by the important work of Bardsley (2008) and List (2007), who showed how giving in dictator games is likely to be affected by experimenter demand effects. Like them, we use dictator games where a subject (the dictator) could give k units of money to another subject at a cost of $a \times k$ units (with a being a transfer price coefficient). Unlike them, we also had a back-to-

¹ One of the main findings of recent experimental research on public good games is the effectiveness of punishment in raising cooperation (e.g., Fehr and Gächter, 2000), particularly with agents that are known to be conditional cooperators (Burlando and Guala, 2005). Masclet et al. (2003) found that agents in public good games were sensitive to even purely social punishment with no monetary implications.

back symmetrical money burning game where the same dictator could destroy k units of money from another subject at the same cost of $a \times k$ units.² Other things being equal (including endowments, transfer price and a one-shot interaction with someone subjects knew they had not been matched with and would not be matched with again in the experiment), other-regarding preferences may explain why a subject may give to a stranger without burning (e.g., altruism), or burn without giving (e.g., spite), or neither give or burn (self-interest). They are not, however, able to explain how a subject may both give and burn under these circumstances. Our proxy for behavioral social desirability is then simply based on the sum of giving and burning in back-to-back dictator and money burning games.³

Section 2 describes the experimental design and results and section 3 briefly concludes.

2. The Experiment

Design. A total of 216 subjects participated to our experiment by coming to two sessions. In the first session they did the social desirability questionnaire. Because of the importance of understanding the questionnaire in its nuances, we aimed to select native English speakers as subjects. In the second session (around 1 or at most 2 weeks later), either they were all invited back to an experimental session (control SDR treatments) or they were invited but sorted depending on whether their social desirability score was high or low (high or low SDR groups). 12 subjects participated to each session 1, 6 to each session 2.

In the second session, subjects carried out five stages, each with a separate player, and subjects knew this. Money was earned in 'experimental points', each worth 1 U.K. penny at the end of the experiment; average payments were 15.55 pounds. The first four stages were pairs of dictator games and money burning games,

² See Zizzo (2003) for a more context-rich multi-player money burning game setup. Needless to say, and as in previous research, the terminology 'money burning' is metaphorical and not literal: money was not physically burnt.

³ More specifically, it will be the average of the corresponding value for each of the two pairs of dictator and money burning games described in section 2. Undoubtedly, this is an *upper limit* to the degree of behavioral social desirability, as it is possible that, for each subject and pair of choices, some burning *or* giving is due, say, to spite or altruism. This needs bearing in mind in interpreting the results.

where subjects played either the role of dictators or that of receivers; receivers did not know the outcome until the end of the experiment. Their endowments were 45 points for the dictator and 145 points for the receiver in stages 1 and 2 and vice versa in stages 3 and 4; this order was counterbalanced across sessions. The transfer price was (1/3), and so dictators could spend up to 15 points to destroy or transfer up to 45 points from/to the receivers.⁴ The back-to-back presentation of each pair of dictator and money burning games was deliberately chosen to make the implied experimenter demand as transparent as possible to the subjects.

Stage 5 was a standard two player linear repeated public good game without or with social punishment technology. Subjects were matched for ten rounds with the same coplayer. For each point given to the public good, they got 0.7 back. Subjects learnt about their earnings after each round. In half of the sessions, we reproduced Masclet et al.'s (2003) instructions on the social punishment technology; as in Masclet et al.'s corresponding treatment, subjects could feed back to their coplayers their disapproval on a numerical scale between 0 and 10 ('disapproval points'), with no financial implications for the coplayers (or for themselves).

Results. Figure 1 displays average giving and burning rates in the burning and dictator games. If we index treatments by the sensitivity to social pressure of their session 2 subjects (Low = 0, Control = 1, High = 2), there is a clear positive correlation between our behavioral social desirability average by session and the social desirability index (Spearman $\rho = 0.341$, p = 0.021). This is non-trivial as sorting occurred based on the Stober (2001) questionnaire scale, and so it suggests that our behavioral social desirability scale indeed captures the social experimenter demand effect. A more direct test of this is by directly correlating Stober social desirability scores and behavioral social desirability average by session: as shown by

⁴ Dictator choices were incentivized. Receivers were asked to make hypothetical dictator and money burning choices.

⁵ This and the other simple statistical tests performed here were run at the session level to address the problem of non independence of observations within session. Note that we did not find more interesting results by looking at behavioral social desirability by endowment, i.e. by looking at whether a subject has 45 points or 145 points in each stage 1-4 game. Also, the same relationship was not found (here or elsewhere) with a behavioral social desirability measure based on hypothetical money burning and giving: real incentives do make a difference in the behavioral tasks.

Figure 2, one does indeed predict the other ($\rho=0.383$, p=0.011). As a further validation of the fact that back-to-back two player dictator and money burning games are mainly driven by social pressure rather than by standard other regarding preferences, Figure 3 verifies whether higher money burning is predicted by higher or by lower dictator giving: if behavior were driven by standard other regarding preferences, the relationship should be negative; instead, we find it is positive, showing that, in sessions where dictators gave more, they also destroyed more ($\rho=0.347$, p=0.038).

Table 1 describes the average public good contribution in each treatment. There is no obvious effect of the social desirability manipulation on behavior, and this is confirmed by session level statistical tests. Table 2 employs tobit random effects regressions controlling for session level effects to investigate the determinants of public good contribution. The Stober questionnaire measure appears statistically significant only in interaction terms, while, as measured with dictators, the behavioral social desirability measure does predict some greater public good contribution, though this effect appears moderated by interaction effects. We do not wish in this note to elaborate *ex post* on the meaningfulness or otherwise of specific interaction effects we made no *ex ante* predictions on; a different paper employs another experiment to test some interaction effects predictions (Fleming and Zizzo, 2009). Table 3 contains further regressions restricted to the punishment treatments, which analyze how, in relation to these treatments, disapproval points affect behavior. There is no evidence that, once one controls for lagged investment decisions, disapproval points in the previous round increase public good contribution, nor does their effect seem to be

⁶ For example, if subjects in a session tend to be altruistic on average, they should give more and burn less; if they tend to be spiteful on average, they should give less and burn more.

⁷ There is also no statistically significant relationship between average disapproval points awarded and our social desirability measures.

⁸ Similar results can be found using tobit random effects regressions controlling for subject level effects. Tobit regressions are required because of the significant number of observations censored at either 0 or 100% contribution.

⁹ We have run a number of robustness checks. For example, the statistical insignificance of the Stober measure in model A in Table 2 holds even if the sample is restricted to dictators as in model B, e.g. the difference in predictive power cannot be explained by the fact that model B employs only a subsample of model A.

moderated by our measures of social pressure. 10 That being said, and as indicated by Tables 1 and 2, the Masclet et al. (2003) prediction of higher contribution under a social punishment technology is supported in one tailed tests (Mann Whitney p = 0.047 at the session level; and p = 0.040 in relation to the punishment treatment dummy of table 2).

Taking Tables 1, 2 and 3 together, the key results appear to be that (a) social pressure, at least as proxied by our measures (or by disapproval points), do not seem to have a large effect on public good contribution, (b) though an average effect can be identified using our behavioral measures, and purely in terms of replicating Masclet et al.'s (2003) result on social punishment technology; (c) there seem to be within-population interaction effects (as per Table 2) worth investigating, if found robust, in future research.

3. Conclusions

Back-to-back giving and money burning occurs and appears driven by social experimental demand. The news with public good contribution appears more positive for experimentalists and more negative for policy makers willing to apply vertical social pressure on agents: public good contribution can only partially be predicted by our measures of social pressure, though at least under certain circumstances social pressure does play a role, as signified by interaction effects, or by the average effect of a social punishment technology. That our measures have limitations does not of course mean that new and better measures may not show a stronger role for social pressure in public settings such as ours. And there may also be other forms of experimenter demand effects – related to task construal as opposed to social pressure – that are not captured by our measures. Obviously, further research is needed.

_

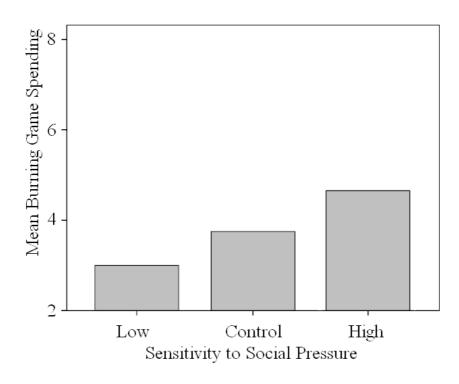
¹⁰ Point coefficients are actually negative, and statistically significant so in model C; however, if one were to control for subject level effects rather than session level effects, the p value would be higher (0.066). Results are otherwise again similar if tobit random effects regressions controlling for subject level effects are used.

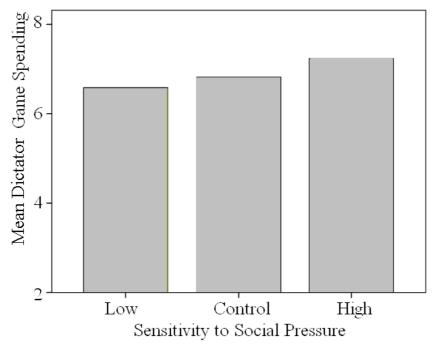
References

- Amos, O.M., 1982. Empirical analysis of motives underlying individual contributions to charity. *Atlantic Economic Journal* 10, 45-52.
- Bardsley, N., 2008. Dictator game giving: Altruism or artefact? *Experimental Economics* 11, 122–133.
- Burlando, R.M., Guala, F., 2005. Overcontribution and decay in public good experiments: a test of the heterogeneous agents hypothesis. *Experimental Economics* 8, 35-54.
- Falk, A., Fischbacher, U., 2006. A theory of reciprocity. Games and Economic Behavior 54, 293-315.
- Fehr, E., Gächter, S., 2000. Fairness and retaliation: The economics of reciprocity. *Journal of Economic Perspectives* 14, 159-181.
- Fehr, E., Schmidt, K., 1999. A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics* 114, 817-868.
- Fleming, P., Zizzo, D.J., 2009. Social desirability, approval and public good contribution, University of East Anglia CBESS Discussion Paper 09-11.
- Larson, M.R., 2000. Social desirability and self-reported weight and height. International Journal of Obesity 24, 663-665.
- List, J.A., 2007. On the interpretation of giving in dictator games. *Journal of Political Economy* 115, 482–493.
- Marquardt, D.W., 1980. You should standardize the predictor variables in your regression models. *Journal of the American Statistical Association* 75, 87-91.
- Masclet, D., Noussair, C., Tucker, S., Villeval, M.C., 2003. Monetary and non-monetary punishment in the voluntary contributions mechanism. *American Economic Review* 93, 366-380.
- Perugini, M., Tan, J.H.W., Zizzo, D.J., 2009. Which is the more predictable gender? Public good contribution and personality. *Economic Issues*, forthcoming.
- Stober, J., 2001. The social desirability scale-17 (SDS17): Convergent validity, discriminant validity, and relationship with age. *European Journal of Psychological Assessment* 17, 222-232.

- Zizzo, D.J., 2003. Money burning and rank egalitarianism with random dictators. *Economics Letters* 81, 263-266.
- Zizzo, D.J., 2009. Experimenter demand effects in economic experiments, *Experimental Economics*, forthcoming.

Figure 1. Mean Money Burning and Dictator Game Spending





Treatments were Low (sorted according to a low score on the Stober, 2001, social desirability measure), Control (unsorted) and High (sorted according to a high score on the Stober measure). Each point spent in the games was equivalent to 3 points burnt from or given to the recipient.

Figure 2. Session Level Correlation between Stober (2001) Social Desirability Score and Behavioral Social Desirability (Bsds) Score

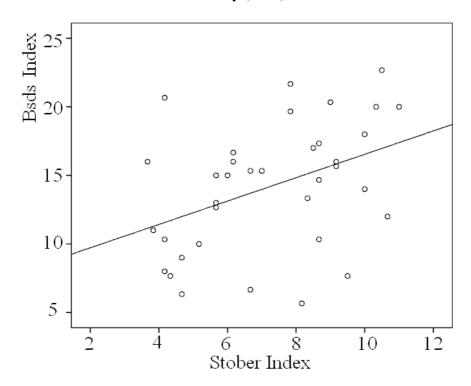


Figure 3. Session Level Correlation between Dictator Game and Money Burning Spending

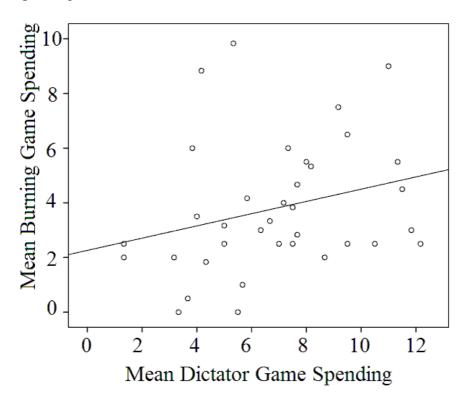


Table 1. Mean Public Good Contribution By Treatment

		Punishment Technology				
		No	Yes	Total		
Social	Low	37.106	52.353	44.729		
Desirability	Control	43.461	46.461	44.961		
	High	38.081	47.508	42.794		
	Total	39.549	48.774	44.162		

Table 2. Tobit Random Effects Regressions on Public Good Contribution

	Model A			Model B		
	β	t	P	β	t	P
PunTr	4.365	1.76	0.079	3.35	1.3	0.195
Stober	-0.05	-0.09	0.927	0.648	2.67	0.008
LInvOwn	0.478	10.85	0	0.463	7.93	0
LInvOth	0.617	12.05	0	0.789	13.35	0
PunTr x LInvOwn	0.048	0.76	0.448	0.054	0.62	0.535
PunTr x LInvOth	0.038	0.59	0.555	0.127	1.46	0.145
PunTr x Stober	-0.952	-1.1	0.27	-0.552	-1.78	0.076
Stober x LInvOwn	-0.046	-3.88	0	-0.014	-2.08	0.038
PunTr x Stober x LInvOwn	0.051	2.56	0.011	0.013	1.55	0.122
Age	0.139	0.52	0.606	0.134	0.44	0.663
Gender	-1.132	-0.44	0.66	2.884	1.04	0.298
NotUK	0.523	0.16	0.87	4.274	1.19	0.233
Economics	-2.64	-0.69	0.49	-1.521	-0.38	0.703
Constant	41.849	17.87	0	40.646	16.72	0
Log likelihood	-6473.609			-3057.406		
Sample size <i>n</i>	1944			972		

PunTr: = 1 if punishment treatment, else 0; SDS = social desirability score according to Stober (2001) in Model A and the behavioral measure in Model B, both centered to avoid multicollinearity problems (Marquardt, 1980); LInvOwn(Oth): centered lagged own (other) contribution; Age: centered aged; Gender: = 1 for men, 0 for women; NotUK: = 1 for non UK subjects, else 0; Economics = 1 for economics students, else 0. Session level effects are controlled in the regressions. Model B is only defined for dictators in stages 1-4.

Table 3. Tobit Random Effects Regressions on Public Good Contribution: Effect of Disapproval Points in Punishment Treatment

	Model C			Model D		
	β	t	P	β	t	P
SDS	-1.236	-2.05	0.04	0.157	0.65	0.517
LInvOwn	0.425	7.58	0	0.411	4.19	0
LInvOth	0.813	16.1	0	0.913	10.21	0
LDisapproval	-0.849	-1.67	0.096	-1.579	-1.87	0.061
LDisapproval x SDS	0.141	0.87	0.382	0.031	0.48	0.631
Age	0.536	1.35	0.176	0.899	1.3	0.192
Gender	-6.416	-2.24	0.025	-4.721	-0.9	0.37
NotUK	-1.236	-0.31	0.759	2.158	0.31	0.757
Economics	-0.066	-0.02	0.988	6.083	0.69	0.492
Constant	49.554	19.98	0	50.269	10.77	0
Log likelihood	-3114.345			-1405.458		
Sample size <i>n</i>	972			486		
Left-censored observations	184			87		
Right-censored observations	224			151		

LDisapproval: centered lagged disapproval points; SDS = social desirability score according to Stober (2001) in Model C and the behavioral measure in Model D, other variables defined as in Table 2. Session level effects are controlled in the regressions. Model D is only defined for dictators in stages 1-4.