Are we capable of being altruistic? Assessing the evidence from experimental economics.

Second Prize – Postgraduate Category

BAZIL A. SANSOM*

1. Introduction

Altruism, defined as *motive*, is intrinsically private and subjective and thus eludes direct observation. For this reason experimental research into other-regarding motives has focused on the elimination of *confounds*, culminating in the double-blind dictator game. This type of experimental work has produced results widely interpreted as strong evidence for altruism. I critically assess this received interpretation of dictator game giving in the light of recent challenges to the established literature. I find grand claims made both to *proof* and *rejection* of evidence for altruism may be overstated, but that recent work does demand a re-evaluation of current theoretical models, and suggest possible directions for a more sophisticated understanding of the psychology of giving and other-regarding behaviour.

* MSc Economics

2. Dictator games as evidence for altruism

The possibility of altruistic motives for giving in the ultimatum game[1] (Güth et al., 1982) is clearly confounded by the possibility that positive offers reflect a strategic self-interested response to fear of rejection (e.g. fear of envy etc.). The dictator game, in which one player - the "dictator" - decides a division, but the other player - unlike the ultimatum game responder - lacks the power to reject, emerged as an apparently elegant solution to this ambiguity (Forsythe, et al., 1994). With strategic considerations eliminated, positive giving in dictator games has been widely interpreted as evidence for an altruistic element in positive ultimatum game offers (Camerer, 2003; Adreoni et. Al., 2007; Fehr and Schmidt, 1999; Bolton and Ockefels, 2000; Ruffle, 1998 etc.). The game became and remains extremely popular among experimental economists. Engel (2011) reports a literature of 129 published contributions testing 616 different treatments with 30 new papers published in 2008 alone. What is more, non-trivial positive giving has been widely replicated. Engel reports a "grand mean" for dictator giving of 28.35% of the endowment pie from his recent meta-study of over 100 experiments - this is significantly more pie than no pie at all - the clear prediction under simple assumptions of rational-self-interest[2]. Is however, elimination of fear-of-rejection, really sufficient to justify claims to proof of altruism by elimination?

[1] In the standard ultimatum game, two players are allocated a sum of money X. One player is assigned the role of first mover or "proposer", and must propose a division of the allocated sum X between themselves and the other player by choosing an amount Y in the interval [0,X] to offer to the other player. The other player, the second mover, or "responder", may *either* accept *or* reject this proposal (Y) (and implied ratio Y/X). If the responder chooses to reject, neither player receives anything. If the responder accepts, the first mover receives (X-Y) and the second mover receives Y. In the standard version this is a one shot interaction (No bargaining). This simple strategic game has a clear game-theoretic Nash-Equilibrium under assumptions of rational self-interest, according to which the responder should accept any Y>0 and the proposer, anticipating this, should offer the smallest possible positive Y. The responder should be indifferent to an offer of zero. Experimental results for the game however, demonstrate systematic divergence from this theoretical bench-mark, with a significant proportion of participants consistently offering positive amounts (see e.g. Camerer 2003 for survey).

[2] This meta study includes it should be noted, some studies which explicitly investigate the impact of social pressure.

It is of course clear that the simple elimination of second mover power to reject does not, of itself, eliminate perhaps the most intractable confound for any motive based definition of altruism: fake altruism. The power of our desire for social approval as motive is well recognised. Attempts to achieve a rigorous test for genuinely other-regarding behaviour have therefore sought to minimize social pressure within the game through the employment of devices such as "double-blind" - guaranteeing dictator-experimenter anonymity as well as between-subject anonymity (introduced by Hoffman et al. 1994, 1996; subsequently Bolton et al., 1998 etc.).

3. Challenging received interpretation of dictator game giving

Some recent work however casts doubt on the meaningfulness of dictator game results, suggesting an insidious distortionary effect from the presence of an experimenter, not satisfactorily mitigated by double-blind type devices (e.g. Bardsley 2008, List 2007, Zizzo & Flemming 2011). These studies question the external validity of dictator games, pointing in particular to a purported *inconsistency between dictator game and real world giving*, as well as to *a high sensitivity to parameters not*, it is claimed, *to be observed in more natural economic settings*. If dictator game results are meaningful, it is asked, why do we not see more anonymous giving outside of the laboratory. The hypothesis that dictator game giving may be an experimental artifact is proposed as an explanation, and new experimental work presented as direct evidence in support of this hypothesis.

3.1 Internal validity of dictator game results?

Bardsley (2008) starts from the theoretical assumption that, given a well-behaved indifference map for other regarding preferences, the introduction of a taking option without any alteration to available giving options, aught not - other things equal - to impact subjects willingness to give. In a two-treatment between-subject modified dictator game with a 1/1 transfer rate and unequal endowments, he finds the introduction of an asymmetrical taking option (with no

alteration to the giving domain), is associated with reduced giving. In another two-treatment between-subjects experiment he ran a similar control treatment[3] and a symmetrical taking game (where the option to give is removed). While generous subjects might be predicted to choose the status quo as the most generous option available, Bardsley finds that only 17% of subjects choose the status-quo in the taking game where 55% of subjects chose to give a positive amount in the standard dictator game - perhaps his most striking result.

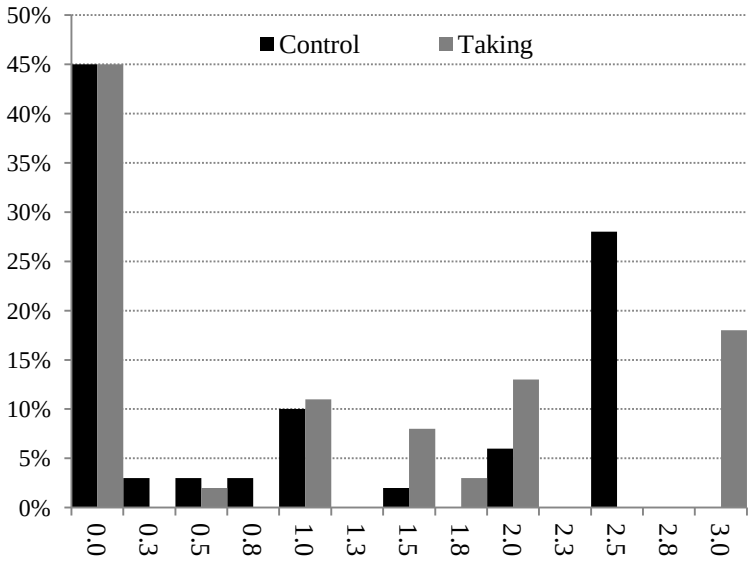
Bardsley claims that this "reversing of generosity between treatments is inconsistent with any theory of dictator game giving which regards underlying other-regarding motivation as causing a desire to share the endowment" (Bardsley 2008, p.128) and himself clearly favours the experimental-demand hypothesis he set out to test. He goes on however, in the same paper, to present the bones of a highly plausible *alternative* explanation for his results, itself *not incompatible* with an updated theory of other-regarding motives: the "perceived kindness of an action may depend on its location in the range available" (Bardsley 2008, p. 129). That is, our perceptions of kindness, or fairness, may be relative/contextual, meaning it may not in fact be the "reversal" of generosity observed, so much as a *subjectivity* or *referent dependence*. This would be consistent with reference dependence in preferences observed elsewhere (e.g. Easterlin, 1995; Clark & Oswald, 1996; Layard, 2003; Kahneman et al., 2000).

In both of Bardsley's above described experiments[4], a majority of participants *did not* choose the most selfish option in either control *or* taking treatments. What is more, it is interesting to note that in both experiments, the results of both treatments describe a remarkably similar bimodal distribution, with the most selfish option the global mode, and a second local maximum around a "fair" division. I have transposed the results from the taking treatment onto those for the control (Fig.1) to illustrate this point. These results are consistent with the possibility of reference dependence in the psychology of other-regarding motives - all be it perhaps equally consistent with a cognitive-experimenter-demand-effect (Zizzo 2010) - thus arguably posing a greater challenge to rationality (challenging e.g. Andreoni & Miller, 2002; less directly Zizzo &

[3] A standard dictator game with 1/1 transfer rate and unequal endowments.
[4] Experiment 2 and 3 in Bardsley 2008.

Oswald, 2001) than to other-regarding motives per-se. Bardsley concedes the need for further empirical work (Bardsley 2008, p.130).

Fig.1: Control and Taking Treatments - Bardsley's experiment-three



Further empirical work is carried out by List (2007)[5]. List largely replicates Bardley's asymmetric taking option experiment[6], but also introduces a game with a *symmetrical* taking option; and experiments with varying the origin of (still unequal) endowments, running one treatment in which endowments are "earned", rather than random allocations as made by Bardsley and in List's own other treatments.

In the game with asymmetric taking option, List, like Bardsley, finds giving reduced in the treatment with an asymmetric taking-option - 35% of subjects making a positive offer in the taking treatment compared to 71% in the control. Interestingly, where List introduces a symmetric taking option, positive giving is almost eliminated, with the most selfish option the modal result at c.40% while a substantial c.30% spike sits right on zero - neither giving nor taking. Thus List's work seems to confirm the main features of Bardsley's results: the most selfish

[5] Note that the work for Bardsley 2008 was known to List as presented Bardsley 2005.
[6] Control treatment and variation on control which includes the option to take a single dollar; unequal endowments; 1/1 transfer rate.

option is the modal choice across treatments, yet a majority of participants fail to choose the most selfish option. Changes to the available choice set seem however, to be associated with dramatically different average giving and distributional outcomes, consistent with a referent dependence in the psychology of other-regarding preferences.

Varying the origin of unequal endowments also yields interesting results: List finds that introducing a "pre-play" session in which participants "*earn"* their endowments, almost eliminates all giving and taking - 66% of subjects choose the status-quo compared to 30% in the control treatment as a result of much reduced "selfish" taking. This result seems highly suggestive of the possible relevance of "just deserts" - and perceptions of normative entitlement - in distribution related decision-making (that is to say, concern for distributional justice). Especially interesting, is that this result implies the possibility that a majority, even of those taking maximally under the taking option treatment in this experiment, are in fact concerned with distributional justice. This raises the possibility that those taking maximally within the game, may consider this behaviour, by some definition, "fair". This would be consistent with the context dependence observed elsewhere for the moral significance of behaviour: *often* in life, "fair", is defined by *the rules of the game* - what is "fair play" in one institutional context, often is not in another. Unfortunately, in the absence of any factorial analysis of these two parameters, we can learn little from List's experiments, about how "just deserts" interact with the apparent sensitivity displayed in distribution related decisions to the range of positive entitlements as defined and enforced by the experimental design.

Overall, List's results seem to help us develop our referent-dependent social-preference hypothesis, without disentangling it much from the alternative experimenter-demand explanation. A common feature of both Bardsley and List's work, is that by employing an incentive structure by which the most selfish choice is to take maximally, it allows space for interpretation of reduced average nominal giving as referent dependence in the resolution of tension between *selfish* and *other-regarding* motives, making evidence for the high level of experimenter-demand suggested, inconclusive.

Zizzo & Flemming (2011) make a provocative contribution to this conundrum, running back-to-back giving and money-burning dictator games with unequal endowments and augmented

transfer rates. Under this incentive structure the neutral point of neither giving nor taking is the clear rational self-interested solution - since not only can a player not *steal* by burning, but actually incurs a cost. Altruists meanwhile, are free to give; while a sufficiently "spiteful" individual may be ready to actually *pay* to reduce the relative value of the other player's endowment. Zizzo & Flemming (2011) expect moreover, that altruistic individuals should burn less, and spiteful individuals give less, while individuals susceptible to experimenter-demand-effects, it is hypothesized, may both give and burn. The authors find a statistically significant positive relationship between session level average giving and average burning in their results (spearman ρ =0.3472 p=0.038) - where each "session" consists of two back-to-back giving/burning games, each game having the inverse direction of inequality in endowments from the other. This is presented as unequivocal evidence for experimenter-demand-effect: social-preferences, we are told "…are not… able to explain how a subject may both give and burn under these circumstances." (Zizzo & Flemming 2007, p.3; see also Zizzo 2011, p.5).

These results certainly seem to challenge received interpretation of dictator game behaviour, but are social-preferences really unable to explain how a subject may both give and burn? Zizzo in an earlier burning experiment, finds that not only are subjects prepared to *pay* to burn other player's endowments, but also, interestingly, that 75% of players are "rank egalitarian" - that is, they burn richer players as much or more than poorer ones, providing support, according to the author, for theories of interdependent preferences that predict agents care about how money is divided among other agents and *not only* about their own relative shares (Zizzo 2002; 2003; Zizzo & Oswald 2001). I propose that given the reversal of unequal endowments Zizzo & Flemming (2011) employ *within* sessions, and the relative size of the two players' endowments, it seems that individuals with any sort of equity based preference for fairness, may be expected both to give and burn - consistent with a session level positive relationship between giving and burning. In fact, given that subjects would have to burn double what they gave in order to achieve a perfectly equal distribution[7], there would appear if anything, to be a structural incentive for any individuals with a preference for fairness, to burn more than they give. Thus it

[7] Note that given the 45/145 and 145/45 starting distribution for endowments, in the two treatments where players have the power to improve the equality of the split, they would have to give 25 and burn 50 (more than their entire endowment) to achieve a perfectly equal distribution.

may *not* be possible to entirely rule out social-preference explanations for the observed positive relationship between giving and burning.

While there seems to be something intuitively wrong with the idea that burning could be other-regarding, if a preference for "fairness" operates *irrespective of the direction of advantage*, this is not consistent with simple self-interest but rather appears pro-social/other-regarding. The positive correlation between giving and burning identified by Zizzo & Flemming (2011) therefore, may, unexpectedly, provide *support* for an other-regarding motives hypothesis[8]. An examination of the relationship between the direction of inequality in endowments and giving and burning behaviour in the study might help to clear up ambiguity on this point[9]. As would running the same experiment with equal endowments and/or 1/1 transfers - there seems no reason to suppose that unequal endowments are intrinsic to the experimenter-demand hypothesis, thus why use augmented-transfers if you are not interested to explore preferences over relative endowments?

3.2 External validity of dictator game results?

These new contributions to the dictator game literature clearly seem to demand we reconsider received interpretation of dictator game giving. It is not clear however that any of the work discussed succeeds in establishing any convincing *internally valid* distinction between experimenter-demand and other-regarding motives. The appeal then, of an experimenter-demand hypothesis over an updated model of other-regarding motives, is that experimenter-demand apparently better accounts for the purported inconsistency between experimental results and "real-world" behaviour. What then, of this question of external validity? Why do we not see more anonymous giving in real-life? I propose that claims to inconsistency between

[8] Although it might be interesting to compare the relationship between giving and burning with those individuals who scored higher on the social-desirability scale stripped out. However the social desirability scale may not provide robust evidence given that a positive relationship between *actual* pro-social behaviour and participants scores (*reported*) should be expected, especially for non-extreme scores.
[9] This data not to my knowledge publicly available nor yet requested of the authors.

dictator game and real-world behaviour stem from a naïve reading of the external relevance of dictator game results.

Bardsley complains "A common concern is that people could always make anonymous donations to random strangers in everyday life, for example by mailing cash to persons sampled from the telephone directory, but few if any choose to do so." (Bardsley 2005, p.1). The personal satisfaction people obtain from being *seen* to be pro-social altruistic beings is hardly in question. The un-natural anonymous design of dictator games is precisely to control for *these* types of motives. Hardly surprising then, if in real-life, given the *choice* people generally take credit for giving. What is more, these sort of complaints miss the fundamental significance of the underlying tension between selfish needs and desires on the one hand; and other-regarding motives on the other, in the context of *unlimited* opportunity for giving. This tricky situation is resolved through personal and social regulating constraints on giving that define essential *agent*-centric limits to moral responsibility/culpability. Where unlimited responsibility would demand *either* discardal of other-regarding considerations, *or* complete *effacement* of self (as per e.g. Nitzsche's view of un-reciprocal altruism as an unacceptable martyrdom (Nitzsche 1887[2007])), these regulating constraints facilitate a managed compromise. Not only that a dictator game is a clearly *limited* one shot interaction making finite moral demands; but also critically the *transparency* of moral responsibility within the procedural relations established by the institutional structure and context of the game, constitute critically defining and *meaningful* features ignored by this type of skepticism regarding external validity.

The work of Dana et al., (2007) on "moral-wiggle-room" showing substantially reduced fairness when the connection between choices and outcomes is obfuscated in dictator games, at once challenges conventional interpretation of giving in dictator games, *and* highlights the importance of moral accountability through a direct investigation of its mechanics, addressing the question of *how* agents "manage" moral responsibility. Their conclusion however - that their results suggest fair behavior may be driven by a desire to appear fair without actually wanting a fair outcome - rests both upon the assumption that other-regarding preferences extend only to other players financial pay-offs not to their feelings; and moreover upon a particularly simplistic

conceptualization of the ego, as unified and coherent. If we allow (1) that participants may be concerned for the social as well as material impact of their choices and (2) that we may not exploit uncertainty or lack of clarity only in the deception of others, but also in self-deception (see Cohen, 2000, for discussion of "denial"); it is not clear that Dana et. al.'s conclusions follow from their results.

Perhaps most importantly though, comments like Bardsley's apparently assume an equivalence between game endowments and "real-world" personal-wealth. However while in dictator games, subjects generally receive arbitrary sums of money, in real-life there is little manna-from-heaven. Mostly we have to work for whatever we get; what is more, wherever unearned privilege does exist, more often than not this is *perceived* or constructed in terms of "just deserts" (the objectively-disadvantaged often coming to *believe* they deserve their lesser outcomes, while the objectively-privileged often come to believe, or persuade themselves, that they are entitled to/deserve their position of advantage (see e.g. Major 1994)). In this context, the suggestion in List's results of the power of perceived "just deserts" in the psychology of normative entitlement to eliminate most giving and taking behaviour irrespective of positive entitlement to take as defined and enforced by the game, or inequality of distribution, clearly suggests another possibly important explanation for why we do not see more "real life" giving. After all, the idea of "just-deserts" as entitlement is enshrined at the ideological heart of "free-market" societies in which everything earns what it is worth and is worth what it earns, and some psychology of entitlement must surely fundamentally underpin the persistence of the unequal outcomes these systems generate.

4. Final remarks

The experiments discussed provide interesting challenges to received interpretation of dictator game giving, but dramatic claims to proof of levels of experimenter-demand effect of a degree that invalidates social-preference interpretations, may be as exaggerated as some of the claims made for proof of altruism. Whilst highlighting the potential significance of cognitive and social

demands placed on the subject by dictator games[10]; the results of these studies may also, unexpectedly, suggest ways in which we might formulate a more sophisticated understanding of other-regarding motives, raising issues - e.g. referent dependence of other regarding preferences; the psychology of "just deserts" in unequal distributional outcomes etc. - of clear relevance to a wide variety of economic settings and problems.

While the sensitivity of dictator games to parameters is referred to critically by some (Bardsley, 2008; Zizzo, 2011); it must be recognised that dictator games are not the only games to display this sensitivity - as Levitt et al., (2007) note, in a wide range of experimental settings, subtle manipulations have been shown to have drastic effects on actions. Where these relationships are systematic, we can learn from them. A single word in an experiment, may be sufficient to entirely alter the game participants are playing - hence defection rates in prisoner dilemma games shift dramatically between "Community" and "Wall Street" games (Ross and Ward, 1996). This highlights a subtle conceptual concern associated with the problem of experimenter-demand: Levitt et al., (2007) note that "…an aspect of the lab over which experimenters have incomplete control is that subjects may not be playing the game that the experimenter intends." (p.163). I suggest that in this sense, a healthy level of experimenter cues may be *essential for interpretability*, just as an experiment must be properly incentivized. But how do we distinguish between experimenter-demand inducing *cues*, and *essential information*, and thus how between experimental *artifact*, and *genuine* other-regarding motives? This may be a question that naïve realism and objectivist inquiry can only approach asymptotically.

[10] See Zizzo, 2010 for discussion.

References

Andreoni, J., and J. Miller (2002) “Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism”, *Econometrica*, Vol. 70, No. 2 (Mar., 2002), pp. 737-753.

Andreoni, J., W. T. Harbaugh, L. Vesterlund (2007) “Altruism in Experiments” Prepared for the *New Palgrave Dictionary of Economics, 2nd Edition,* 2007

Bardsley, N.,. (2005) “Dictator Game Giving: Altruism or Artefact?”, ESRC National Centre for Research Methods NCRM Working Paper Series.

Bardsley, N., (2008) “Dictator game giving: altruism or artefact?”, *Experimental Economics* 2008 11, 122–133.

Bolton, G. E., R. Zwick, and E. Katok (1998) “Dictator game giving: Rules of fairness versus acts of kindness,” *International Journal of Game Theory*, 27, 269–299.

Bolton, G. E., and A. Ockenfels (2000) “ERC: A Theory of Equity, Reciprocity, and Competition.”, *American Economic Review*, 90 (March): 166–93.

Camerer, C. F. (2003) “*Behavioral game theory. Experiments in strategic interaction*.” New York: Sage.

Clark, A. and A. Oswald (1996) "Satisfaction and comparison income", *Journal of Public Economics*, 1996, 61, pp. 359-381.

Cohen, Stanley (2000), “States of Denial: Knowing about Atrocities and Suffering”, Polity Press (14 Dec 2000).

Dana, J., R. A. Weber, and J. Kuang (2007), "Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness", *Economic Theory*, Vol. 33, No. 1, pp. 67-80.

Easterlin, R. (1995) "Will raising the incomes of all increase the happiness of all?", *Journal of Economic Behaviour and Organization*, 27, pp. 35-47.

Engel, C. (2011) “Dictator games: A meta study.” *Experimental Economics*, Volume 14, Number 4, 583-610.

Fehr, E., and K. M. Schmidt (1999) “A Theory of Fairness, Competition, and Cooperation.” *Quarterly Journal of Economics* 114 (August): 817–68.

Forsythe, R., J. L. Horowitz, N. E. Savin and M. Sefton (1994) “Fairness in Simple Bargaining Experiments.” *Games and Economic Behavior, 6,* 347-369.

Güth, W., Schmittberger, and Schwarze (1982) "An Experimental Analysis of Ultimatum Bargaining." *Journal of Economic Behavior and Organization* 3 (4): 367–388.

Hoffman, E., K. McCabe, and V. Smith (1996) "Social Distance and Other-Regarding Behav- ior in Dictator Games," *American Economic Review*, 86, 563–660. [1608]

Hoffman, E., K. McCabe, K. Shachat, and V. Smith (1994) ''Preferences, Property Rights, and Anonymity in Bargaining Games,'' *Games and Economic Behavior*. 7, 346380.

Kahneman, D. and A. Tversky (2000) "Choices, Values, and Frames", Princeton University Press.

Layard, R. (2003) "What causes happiness? Rethinking public economics", Lionel Robbins Memorial Lectures, available at: http://cep.lse.ac.uk/events/lectures/.

Levitt, S. D., and J. A. List (2007) "What Do Laboratory Experiments Measuring Social Preferences Reveal about the Real World?" *Journal of Economic Perspectives,* 21(2): 153–174.

List, J.A., (2007) "On the interpretation of giving in dictator games." *Journal of Political Economy* 115 (2007), 482–493.

Major, B., (1994) "From social inequality to personal entitlement: the role of social comparisons, legitimacy appraisals, and group membership.", *Advances in Experimental Social Psychology*, Volume 26, 1994, Pages 293–355.

Nietzsche, F. W., (1887) "On the genealogy of morality", Cambridge : Cambridge University Press 2007.

Ross, L., and A. Ward (1995) "Naive Realism: Implications for Social Conflict and Misunderstanding", Stanford Center on Conflict and Negotiation: Working Paper No. 48

Ruffle, B.J. (1998), "More is better, but fair is fair: Tipping in dictator and ultimatum games", *Games and Economic Behavior*, vol. 23, pp. 247-265.

Zizzo, D. J., and A.J. Oswald (2001) "Are People Willing to Pay to Reduce Others' Incomes?", accessed: http://www.andrewoswald.com/docs/finaljuly13paris.pdf

Zizzo, D. J., (2002) "Fear the evil eye", Oxford Department of Economics Discussion papers, No.91, March 2002.

Zizzo, D.J., (2003) "Money burning and rank egalitarianism with random dictators.", *Economics Letters* 81 (2003), 263–266.

Zizzo, D. J. and P. Fleming, (2007) "Can experimental measures of sensitivity to social pressure predict public good contribution?", CBESS Discussion Paper 10-03.

Zizzo, D.J., (2011) “Do Dictator Games Measure Altruism?”, forthcoming *Handbook on the Economics of Philanthropy, Reciprocity and Social Enterprise* edited by Luigino Bruni and Stefano Zamagni and to be published by Edward Elgar. August 2011.

Zizzo, D. J. and P. Fleming, (2011) “Can experimental measures of sensitivity to social pressure predict public good contribution?”, *Economics Letters* 111 (2011) 239–242.