

**The debate around lab dictator games:
Should the baby and the bath water be thrown out?**

Nathalie Etchart-Vincent*

Short Abstract: A still ongoing debate opposes those who consider that valuable lessons (regarding pro-social behaviour) can be drawn from lab experiments and dictator games, and those who refer to (natural) field experiments to highlight the lack of external validity of lab experiments as regards giving behaviour. In this paper, we discuss several oft-mentioned arguments to try and show that lab experiments and dictators are still worth using to investigate pro-social behaviour, and also that we had better consider lab and field experiments as complementary, rather than competing, empirical tools.

JEL: C93, D64, D83

Keywords: dictator game, giving behaviour, external validity, lab experiment, natural field experiment, methodology

* Centre National de la Recherche Scientifique (CNRS) & Centre d'économie de la Sorbonne (CES).

Correspondence: Nathalie Etchart-Vincent, CES, Ecole Normale Supérieure de Cachan, 61, Avenue du Président Wilson, 94235 Cachan Cedex, France; Tel: (+33) 1 47 40 74 59; Email: nathalie.etchart-vincent@ens-cachan.fr

Abstract: The status of the dictator game has changed in the recent years. Due to its simplicity and transparency, it was first considered as a wonderful device for properly testing the main behavioural prediction of standard game theory that people are self-interested. First, the simplicity of the dictator game makes it unlikely that the subjects may not understand the logic of the game, so any deviation from the theoretically predicted behaviour (i.e. the existence of the positive transfer) should not be attributed to irrationality (Camerer and Thaler, 1995; Guala and Mittone, 2010). Second, contrary to what happens in the ultimatum game, where the fear of retaliation is likely to affect the first mover's decision and confound the interpretation of her behaviour, any positive transfer can be unambiguously considered as a "pure" pro-social behaviour (Kahneman et al., 1986; Forsythe et al., 1994). On average across experimental studies, about 60% of dictators appear to give a positive transfer and their transfer is about 20% of their endowment (Camerer, 2001). The undoubted prevalence of pro-social behaviour in the dictator has made the dictator a figurehead of behavioural economics, leading scholars to focus on its motives as well as to enrich standard theory to make it more descriptive.

But in the most recent years, several influent and respected scholars have drawn attention to what they consider to be a huge discrepancy between the high prevalence of other-regarding behaviour in the lab and the prevalence of selfish, even dishonest, behaviour in real life. The dictator game has been shown not to be a proper tool to investigate real behaviour, and especially real *giving* behaviour, with results being too sensitive to the very features of the experimental design, such as the choice set (Bardsley, 2008; List, 2007). In their 2009 best-seller book entitled *SuperFreakonomics*, Steven Levitt and Stephen Dubner extensively describe the famous criminal case of Kitty Genovese's murder in 1964 (a murder that 38 neighbours witnessed through their window, without either moving or calling for help) to vigorously emphasize that people can be particularly selfish, and they further refer to John List's work (and especially to his field study on the baseball cards exchange market) to claim that, when making decisions in the field while ignoring that their behaviour is scrutinized, people do actually not exhibit the same altruistic behaviour as in dictator games in the lab. Simplification for a better understanding could be the reason for the entrenched position taken in this book. But the same clear-cut message emerges from List (2006a), who even shows that deviations from theoretical predictions, that recurrently occur in the lab, do actually disappear in the field, suggesting that the lab is both unable to describe actual behaviour and to accurately test theories. Since 2006, John List and coauthors have published several important methodologically-oriented papers, aiming or allowing to compare the performance of the lab and the field, especially in the domain of pro-social preferences (Levitt, 2007; Levitt and List, 2007b; List, 2006b, 2008a, 2009, 2011; see also Winking and Mizer, 2013). Reading them carefully shows that their position may be somewhat more subtle, but two recurrent findings emerge from their many results, that can be easily leveraged against both the resort to lab experiments in general and the use of dictator games (as a symbolic lab experimental tool) in particular : 1) the allegedly high level of giving observed in lab dictator games is actually an

experimental artefact (due to lab unrealistic features and constraints) resulting in poor *external validity*, that is in a basic inability to inform us about people's genuine behaviour in real interactive decision settings, 2) this poor external validity implies the methodological superiority of field experiments when a descriptive and prescriptive, or even maybe theoretical, purpose is at stake.

What we intend to do here is try to sort out, among the set of arguments levelled against the use of the dictator game in the lab, admissible pieces of criticism from more questionable arguments. Our main point will be to show that, as regards the investigation of pro-social behaviour, a number of misconceptions and over-generalizations should be revised:

- 1) The name "field experiment" actually includes a wide range of studies, from the artefactual field study (which is very close to a standard lab experiment) to the natural field study (which lets subjects participate in the experiment without being aware of it, and offers a close proxy to real life) (Harrison and List, 2004). The outcome of the comparison between behaviour in a field study and behaviour in a lab study will obviously depend on "how much field there is" in the field study, preventing us from any clear-cut conclusion *a priori*. However, most of the time, comparisons are made between lab experiments and *natural* field experiments, which provide both a methodological benchmark (lab vs. field) and a descriptive benchmark (lab vs. real life). Most pieces of criticism against lab experiments actually spring from such comparisons (List, 2009).
- 2) The name "lab experiment" actually includes a wide range of studies, allowing many variations in the experimental design that may strongly affect behaviour (for instance, the degree of anonymity or social distance between allocator and recipient may vary; the initial endowment may fall from heaven or be earned by the subject, see Cappelen et al., 2013; see also Engel, 2010 for a meta-study). Therefore, that kind of uniformly pro-social behaviour that average figures suggest (with 60% of giving subjects and about 20% of the initial endowment transferred) is only a statistical artefact, concealing the reality of huge heterogeneity across studies. For instance, behaviour in a dictator appears to strongly depend on the choice set (which may or not include the possibility of taking money, see Bardsley, 2008; List, 2007; Korenok et al., 2014; but see Grossman and Eckel, 2012, for an opposite result suggesting that the previous result could only be an experimental artefact). This does not imply that dictator games are not valuable, it simply reminds us that the word "dictator game" actually refers to many different decision settings.
- 3) Some decision contexts or research questions may exist for which field studies appear to be more appropriate than lab studies, but this does not imply an outright superiority of the former over the latter. More generally, lab and field experiments should be seen as complementary, rather than competing, devices. This complementarity may take three forms. First, depending on the research question (that is, not universally), the best strategy allowing to answer it may be either a lab or a field experiment. For instance, some changes in the decision setting (e.g. a blind vs.

double-blind procedure) might be easier to implement in the lab than in the field, while other (e.g. high stakes) might be easier to implement in the field. Second, findings from lab and field experiments should be considered as elements of a single corpus that should be examined as a whole. When the same question has been investigated both in the lab and in the field (e.g. the influence of anonymity on giving behaviour), comparison should be made between the studies, not to determine which strategy is the best, but to identify, in a constructive spirit, what drives differences in apparently similar decision settings. This comprehensive approach may help to understand which features of the decision setting strongly affect decision making, and which do not. Thirdly, it may be useful to run both lab and field experiments in a same project, in order to investigate the influence that a given ingredient may have on the correlation between behaviour in the lab and in the field. For instance, using a public goods game, Englmaier and Gebhardt (2012) show that the only ingredient that strongly affects the correlation between behaviour in the lab and the field is the strategic component of the game (i.e. the incentive to free ride).

- 4) One should be cautious when claiming that behaviour in the lab is pro-social while behaviour in the field is selfish. First, as shown before, behaviour in the lab is not uniformly pro-social (see Hoffman et al., 1994, 1996 for comprehensive studies). Second, behaviour in the field is not so uniformly selfish (or even dishonest) as suggested by List (2006a)'s sport cards exchange example. The huge amount of charitable giving and volunteering observed in the field shows that people may be strongly pro-social in the field, too. Third, it is crucial to disentangle the behaviour and its motives. Behaviour can be apparently pro-social (i.e. people may be more willing to give than theoretically predicted) without their preferences being necessarily altruistic (Camerer and Thaler, 1995): giving may result from self-interested motives. What seems to be at stake is rather the *interpretation* of behaviour (as resulting from genuinely altruistic preferences) than behaviour itself (List, 2007). Maybe new theories are needed to account for unselfish behaviour without resorting to other-regarding preferences (Korenok et al., 2014). That said, this may help solve the apparent contradiction between the selfish behaviour in the field and (only apparently unselfish) behaviour in the lab. Fourth, the ongoing debate clearly suggests that one should avoid extrapolating the findings obtained in a given decision setting (be it in the lab or in the field) to another decision setting without caution. Additionally, one should not reproach lab experimenters for obtaining giving rates that obviously over-estimate real giving rates, as soon as the aim of the experiment is to show the existence of the giving behaviour phenomenon (and its *relative* prevalence depending on the decision setting) rather than to get a reliable quantification (an *absolute* value) of this giving behaviour. More generally, it may be useful to stop focusing on giving rates and accept that comparisons between the lab and the field should be made at a qualitative rather than quantitative level. If some differences between the lab and the field are irreducible, quantitative comparison may just be meaningless, while qualitative evaluation may help calm down the ongoing debate (Kessler and Vesterlund, 2014).