

Field experiments in the developed world: an introduction

John A. List* and Robert Metcalfe**

Abstract Field experiments represent a relatively new area in economics to understand the causal links from one variable to another. They have been used by academics to help answer interesting and policy-relevant questions in the developed world relating to educational attainment, tax avoidance, consumer finance, negative externalities, charitable giving, and labour market contracts. In this paper we bring together the key ideas behind the different variants of field experiments, how field experiments have been used to test theory, their limitations, and the new areas currently being opened up by field experiments.

Keywords: field experiments, causality, public policy

JEL classification: C9

I. Introduction—why experiment in the field?

This issue of the *Oxford Review of Economic Policy* focuses on the use of field experiments in economics. Our overall aim is to provide the reader with more information about field experiments: what they are; how they are being developed in a number of different areas; and their advantages and limitations as a research method for economic and social sciences.

The experimental approach is widely used, for example, in the medical sciences. The approach of conducting field experiments in economics refers, quite simply, to the application of experimental methods in the ‘field’—namely in the ‘real world’ or with actual market participants, rather than in the laboratory with student subjects. Field experiments in economics are relatively new, yet they have become one of the fastest growing and ‘fashionable’ methodologies in economics and the social sciences in recent years. There are several reasons for this. One is the increasing emphasis among policy-makers on ‘evidence-based policy’: field experiments offer the prospect of determining

* University of Chicago, e-mail: jlist@uchicago.edu

** University of Chicago, e-mail: metcalfe@uchicago.edu

We wish to thank Chris Allsopp and Cameron Hepburn for their excellent editorial comments and support, the OxREP editorial board for initiating this issue, Alison Gomm for her excellent administrative support, and Greer Gosnell and Justin Holz for their research support. We would also like to thank the contributors to this edition, and the referees that have provided excellent timely reports on the contributions. doi:10.1093/oxrep/grv005

© The Authors 2014. Published by Oxford University Press.

For permissions please e-mail: journals.permissions@oup.com

what ‘works’, and what does not work. More generally, field experiments are useful in assessing causality, which is a requirement if hypotheses are to be properly tested. Empirical verification requires that the decision-makers under study operate in a real world context, subject to actual stakes, incentives, constraints, and contexts.^{1,2} This is not true in laboratory-type experiments.

Field experiments, if well designed, directly tackle the pervasive counterfactual problem—by which we mean establishing what would have happened in the absence of the intervention (or treatment) under investigation. By creating correct counterfactuals, unobservable influences can be balanced, allowing the researcher to overcome confounding issues that have plagued other empirical approaches attempting to go beyond correlations to establish causality. Thus an important advantage of field experiments is their ability to provide the researcher with causal relationships in naturally occurring settings. Field experiments can control the variable of interest, but cannot necessarily control the changing context of the environment. Field experiments have been used primarily in microeconomics research because the researcher is often in a position to control the variable of interest. This is less likely to be the case in macroeconomics. For example, it is difficult to conduct field experiments that examine the impact of changing tax rates, interest rates, or the money supply. These, however, could be fruitful areas for future research.

An advantage of field experiments is that they lend themselves easily to a routinized programme of evidence-based research (as, for example, in medicine) in a way that other empirical methods for causal inference do not (at least not yet). Equally simple approaches that could be easily routinized—such as propensity score matching—require a different, and often more restrictive, set of assumptions to go beyond correlations. In the policy world, correlations incorrectly viewed as causal can be very detrimental. Some have argued that field experiments can be the ‘gold standard’ in establishing more-effective public policies. Others remain sceptical, sometimes on grounds of cost.

Finally, field experiments offer more transparent identification than many of the other empirical methods. Their simplicity does, indeed, hide more complex statistical and econometric issues in some cases, but the ease of conveying straightforward and comprehensible information to academic peers, policy-makers, CEOs, etc., strengthens the impact of field experiments.

There are a number of areas in economics that have attempted to use field experiments. Many of those areas are represented in this issue of the *Oxford Review of Economic Policy*, including education, environmental conservation, taxation, charitable giving, personal finance, and labour supply. Our focus here is on the developed world, though many lessons from research in this issue are relevant also to the developing world. What runs throughout these papers is that field experiments have been used to address some important public and economic policy challenges. As we have noted, the area of field experiments is relatively young, and developing rapidly.

¹ There is currently a rich debate about the differences and similarities between the laboratory and the field in economic research. The interested reader should consult Levitt and List (2007a,b, 2009), Camerer (2011), and Al-Ubaydli and List (2013).

² Field experiments can be artefactual, framed, or natural. We will go into the elements of these three ‘types’ of experiments below, but the central tenet that runs through them is the construction of the best possible counterfactual to understand relevant treatment effects.

Given we are both users and advocates of field experiments, this introductory article has a positive and optimistic slant on the area. We recognize, however, that there are a number of methodological and even philosophical issues relating to field experiments that have not been fully resolved. Moreover, as a policy tool, they have practical issues such as cost as well as ethical issues (such as arise in medicine). There has, indeed, been some criticism on the field experimental approach, especially when it comes to drawing inferences from experiments for theory and external validity (Deaton, 2010). By external validity, we mean how the results from the field experiment generalize to other contexts and populations. Cartwright (2012) places a spotlight on the external validity of a randomized controlled trial (RCT). Indeed, RCTs can have problems with external validity, but RCTs are only one part of the field experimenter's toolkit. We would argue that natural field experiments have greater external validity than classic RCTs. We fully agree that the contextual factors that have an impact on field experiments need to be fully explored. We consider some of the criticisms further below.

Section II examines the definitional and conceptual issues, including both supportive and critical perspectives in, we hope, a balanced review. Section III is a more detailed account of the articles in this issue of the *Review*, which sets out the overall issue structure, and elaborates on the individual articles and the common themes between them. The concluding section discusses what can and cannot be gleaned from the articles, what field experiments have contributed and might contribute in the future to economic policy-making, and where the weaker and more fruitful areas for future research and application might lie.

II. What is a field experiment?

We have already noted that field experiments lie on a spectrum between laboratory experiments, where all conditions and interventions are controlled (controlled data) and studies based on naturally occurring data. Figure 1 (taken from List (2006); see also Levitt and List (2009)) illustrates this, putting (our assessment of) a number of well-known empirical techniques, from left to right, along a line from laboratory experiments (where all data is controlled), to those on the far right (where all data are naturally occurring). The various types of field experiments (see below) lie 'on the bridge' between the laboratory and the 'real world' of naturally occurring data.

Essentially, field experiments involve a mixture of naturally occurring data and the experimental method. Since the purpose of experimentation is to help to determine causal

Figure 1: Spectrum of empirical methods



Notes: Lab: lab experiment; AFE: artefactual field experiment; FFE: framed field experiment; NFE: natural field experiment; NE: natural experiment; PSM: propensity score estimation; IV: instrumental variables estimation; STR: structural modelling.

Source: List (2006).

relationships of interest to the researcher, why not go the whole way, and rely on laboratory-type experiments? We would stress two reasons. The first is that it may, simply, not be possible for practical reasons. The second is that, for sentient beings, the context is likely to be very important. For example, experiments in the lab, even if carefully controlled (for example with, double-blind techniques³) may give rise to results that are not externally valid because they lack the incentives, constraints, and the context of the choice environment. At the very least, there is a case for checking laboratory-type results in the field (see the paper by Jasper and Samek in this issue). We would go further: as a research method it widens the set of tools available, both to economists and to policy-makers. As such, field experiments may widen the set of economic and policy issues that are susceptible to empirical testing. Our assumption is that, as economists, we are more interested in learning about understanding behaviour in natural settings than in non-natural settings.

It is also useful to delineate what field experiments are not—since there are many prevalent misconceptions. First, field experiments are not behavioural economics. The former is a research method, the latter is a field or a body of research insights, some of which have been unearthed using field experimental methods. Second, field experiments are not pilot studies. A field experiment can be done on the national rollout of a policy, and, equally, pilots are not necessarily field experiments. Third, field experiments are not RCTs. Randomization is often used in various settings, such as the laboratory, and all field experiments do not necessarily have randomization (though many do).

Returning to Figure 1, it will be noticed that we have included a number of empirical techniques at the far right—i.e. techniques which have been developed to deal with identification problems which arise in using naturally occurring data. The goal of any evaluation method is to construct the proper counterfactual. In a simple example, where the problem is to identify the effects of a particular ‘treatment’ or intervention, define y_{i1} as the outcome for person i with treatment, y_{i0} as the outcome for person i without treatment. The treatment effect for person i can then be measured as $\tau_i = y_{i1} - y_{i0}$. The major problem, however, is that person i is often not observed in both states—this is the problem of the missing counterfactual. In the laboratory, the researcher can construct appropriate counterfactuals, for example by randomization.

One technique for identification in natural experiments results from a difference-in-difference regression model: $Y_{it} = X_{it}\beta + \tau T_{it} + \eta_{it}$, where i indexes the unit of observation, t indexes years, Y_{it} is the outcome, X_{it} is a vector of controls, T_{it} is a binary treatment variable, $\eta_{it} = \alpha_i + \lambda_t + \varepsilon_{it}$, and τ is measured by comparing the difference in outcomes before and after for the treated group with the before and after outcomes for the non-treated group.⁴ A major identifying assumption in this case is that there are no time-varying, unit-specific shocks to the outcome variable that are correlated with T_{it} , and that selection into treatment is independent of the temporary individual-specific effect.

Useful alternatives include the method of propensity score matching (PSM) developed in Rosenbaum and Rubin (1983). Again, if both states of the world were observable, the average treatment effect, τ , would equal $\bar{y}_1 - \bar{y}_0$. However, given that only y_1

³ Where neither the experimenter nor the participants know whether they are in the treatment group or not.

⁴ Note that in this formulation the analyst is assuming a common treatment effect, τ , rather than a heterogeneous treatment effect, τ_i . Use of a random coefficients econometric model in this framework permits estimation of heterogeneous treatment effects.

or y_0 is observed for each observation, unless assignment into the treatment group is random, then, generally, $\tau \neq \bar{y}_1 - \bar{y}_0$, and the estimate of the treatment effect would be ‘confounded’. The solution advocated by Rosenbaum and Rubin (1983) is essentially a statistical matching technique to make sure that, as far as possible, the control group is an appropriate counterfactual for the treated group in the sense that (conditional on their procedures) the non-treated outcomes are what the treated outcomes would have been if they had not been treated, thus mimicking a randomized experiment.

Other well-known methods of measurement include the use of instrumental variables (see Rosenzweig and Wolpin, 2000) and structural modelling. The assumptions behind these approaches are well documented and are not discussed further here (but see Blundell and Costa Dias (2002) for a useful review). Between laboratory experiments and approaches with naturally occurring data in Figure 1 are the various types of field experiments. We now turn to a more detailed discussion of controlled field experimentation.

Clearly, field experiments can be used in all sorts of different contexts and the research design will depend on the context. Harrison and List (2004) propose six factors that can be used to determine the field context of an experiment: (i) the nature of the subject pool; (ii) the nature of the information that the subjects bring to the task; (iii) the nature of the commodity; (iv) the nature of the task or trading rules applied; (v) the nature of the stakes; and (vi) the environment in which the subjects operate. The appropriate research design will depend on all these.

It is useful, however, to categorize field experiments into three types: (i) an artefactual field experiment (AFE); (ii) a framed field experiment (FFE); (iii) a natural field experiment (NFE).

AFEs use non-standard subject pools directly related to the researcher’s research question. They have primarily been used to compare student behaviour to more relevant populations (e.g. List and Mason, 2011), or to incorporate important population-specific characteristics into the research design (e.g. Harbaugh *et al.*, 2002; Gneezy *et al.*, 2009). FFEs move a step closer to the naturally-occurring setting by introducing a natural context, though they may still suffer from selection bias and experimenter effects. Finally, NFEs overlay randomized incentives or contextual factors on subjects in their natural environment where they do not know that they are participating in an experiment. These experiments are particularly useful when selection into the experiment or demand effects otherwise might confound identification and proper interpretation (e.g. DellaVigna *et al.*, 2012; Bertrand and Mullainathan, 2004).

Many papers have reviewed the literature on the design of field experiments (see Harrison and List, 2004; List, 2006; Reiley and List, 2007; Levitt and List, 2009). The increased use of field experiments reflects a long running trend in applied microeconomics to identify causal effects. List and Rasul (2011) argue that field experiments have become more prevalent because of an acceleration of three long-standing trends in how applied economic research is conducted: (i) the increased use of research designs to uncover credible causal effects; (ii) the increased propensity to engage in primary data collection; and (iii) the formation of ever-closer academic interactions with practitioners and policy-makers more generally.

More recently there has been an increase in field experiments designed to estimate structural parameters in fully specified economic models. Unlike reduced-form work, structural field experiments allow the research to go beyond the estimation of treatment

effects and estimate full distributions along with social welfare effects (see, for example, DellaVigna *et al.*, 2012, 2014). Moreover, if well designed, these experiments can reduce the number of identifying assumptions needed in naturally occurring data by explicitly creating exogenous variation in the subjects' environment.

There are many areas in which field experiments could potentially be helpful, but which remain unexplored. One constraint is that the pros (and cons) of this relatively young set of techniques may not be well understood by organizations, such as national governments and other agencies. Moreover, as far as the researchers are concerned, conducting field experiments with organizations requires a set of skills that might not be accessible to all economists. This skillset includes the ability to present and communicate ideas to lay individuals, the patience to work to others' timescales, and the understanding of the value of such experiments for organizations. List (2011a) documents a list of recommendations to help economists carry out field experiments with partners ranging from government agencies to Fortune 500 companies to non-profit organizations.

III. This issue

Beyond their use in economics, field experiments are increasingly used by organizations, both private and public. We have therefore chosen topics in this issue that are important to policy-makers and to many private organizations in the developed world.⁵ We start off with the large topics of human capital (Sadoff), environmental externalities (Price), tax evasion (Hallsworth), public goods provision (Jasper and Samek), consumer finance (Lusardi *et al.*), and labour supply (Levitt and Neckerman). These are the areas where field experiments have made a significant contribution. We then include two rather different papers. One focuses on how best to give the right sets of tools to policy-makers and organizations to conduct more field experiments (Dolan and Galizzi). The other takes a critical perspective, focusing on the limitations of field experimentation (Harrison). It is hoped that these latter two papers will be helpful to understanding how to promote appropriate experiments in context where they are likely to be most useful. This issue is the first collection of papers in this area with such topical variety.

We start with a topic of extreme interest within economics: how do you improve human capital and what are its impacts on wages? More specifically, Sally Sadoff (2014, this issue) addresses the issue of what policy interventions, designed to improve educational attainment have been tested in the field. The amount of schooling per individuals over the last 80 years has been dramatic. For instance, while less than 25 per cent of Americans born in 1930 attended college, over 60 per cent of those born in 1970 did so (Goldin and Katz, 2008). This radical change, along with the increased saliency of the debate over schooling, led to increased thinking about causality in the early 1990s, and was a catalyst for much econometric research in causality (see Card and Krueger, 1992). The issue of how best to improve people's human capital is as important as it was 20 years ago when the causality debate started. Some educational policy-makers are starting to understand that field experiments are an important source of evidence that

⁵ The two largest omissions in this issue are the areas of health and criminal behaviour. There is research in these two areas and the interested reader should consult List and Samek (2014) and Downs *et al.* (2009) for examples.

could be used more often to help inform policy-makers (e.g. the Education Endowment Foundation in the UK).

We then move to thinking about a large negative externality in our society—air pollution. Michael Price (2014, this issue) reviews the literature on the experiments that have been conducted to understand how to address the negative environmental externalities from our consumption. There has recently been an increase in the number of studies designed to test economic and behavioural theory in energy markets—in part due to the large amounts of data available on both the demand and supply side. Given that climate change is driving policy-makers to think more clearly about how their policies affect the environment (Helm and Hepburn, 2009), field experiments can be especially useful for policy-makers in this area.

Establishing causality has been a large area within labour economics, especially since the Card and Krueger (1992) difference-in-difference analysis of the minimum wage. Trying to implement experiments with firms to understand the basic theories in labour economics has been difficult. This is where Steven Levitt and Susanne Neckermann's paper (2014, this issue) is directed. They review some of the field experiments that have been done on changing different structures of wages for employees, or designing different work environments. They document some fascinating work in this area but note that we actually know very little about some fundamental parts of businesses, such as the impact of training and recruiting schemes, team structure, and contract design on productivity and revenue.

Another area where field experiments have been widely used is payment compliance. In particular, tax evasion and compliance has been the focus of much historical economic research, and applying experimental tools is helping to identify causes and solutions to the problem. Michael Hallsworth (2014, this issue) reviews the experimental literature on payment evasion and compliance. This work is directly relevant to many different types of organizations around the world, as getting people to declare information or pay a fee is an important part of any business or organization.

In addition to growing experimental research on compulsory payments, there is an increasing literature on voluntary payment for public goods. Cynthia Jasper and Anya Samek (2014, this issue) review the literature on charitable giving, philanthropy, and field experiments. They then provide evidence from their own field experiment in this area. Most of this literature has focused on differentiating between various different possible motivations for charitable giving. For example, these studies examine whether variations in price or third-party contributions lead to perfect crowding-out as predicted by altruism, or imperfect crowding-out, as predicted by other models such as 'warm glow' (Andreoni, 1990), status (Ball *et al.*, 2001; Huberman *et al.*, 2004), or social pressure (Ariely *et al.*, 2009; DellaVigna *et al.*, 2012). The debate about the relative importance of altruism and 'warm glow' in predicting human behaviour is an emerging research area within economics (see List, 2011b).

These field experiments have also provided insights into the limitations of laboratory studies in these areas. A meta-analysis of over a hundred experiments using dictator games⁶ found that, on average, dictators send 28.35 per cent of their endowments, 63.89

⁶ In this class of 'game', a 'dictator' elects to share a proportion of windfall with a recipient who is entirely passive. In contrast, in an ultimatum game, the recipient can choose to reject the division (in which case neither player gets any part of the windfall).

per cent of dictators share some portion of their endowment, and 16.74 per cent choose an equal split (Engel, 2011). However, Winking and Mizer (2013) compared giving rates in dictator-type games between a laboratory and natural field experiment. They found that no participants gave money in the natural field setting, while giving rates under laboratory conditions were similar to those in past studies. Levitt and List (2007a) and List (2007) provide frameworks for thinking about these differences. They argue that the scrutiny of the lab causes an individual's choices to depend on factors other than the incentives explicitly laid out by the experimenter. These context effects preclude the experimenter from cleanly identifying a participant's preferences.

Another direction for field experimental work has been in the area of personal finance. Across many countries, individuals are being given greater autonomy over their savings and investments. Despite this increase autonomy, the basic level of economic and financial education is low (Hung *et al.*, 2009). Aileen Heinberg, Angela Hung, Arie Kapteyn, Annamaria Lusardi, Anya Samek, and Joanne Yoong, in their article (2014, this issue), present an overall review of this literature, and then describe their innovative 'Five Steps' programme, that covers basic financial planning and education. We see using field experiments to address issues relating to economic and financial education becoming far more widespread in the future, as people will probably have even more autonomy going forward.

In the penultimate paper, Paul Dolan and Matteo Galizzi (2014, this issue) focus on ways to persuade policy-makers and organizations to use and apply field experiments. While economists are beginning to understand the uses and applications of field experiments, the techniques are not well understood among many of those who have the opportunity to implement these experiments in their organizations. Improved understanding is important, given the amounts of money organizations spend on consumer insight to predict how to change human behaviour. The breadth and depth of consumer insight in organizations is vast, yet this methodology is hardly tested in reality (see Graves, 2010).

The final paper casts a critical lens on field experiments and how economists should best use them. Glenn Harrison has been an active advocate of field experimentation, but, in this issue, he examines some important limitations of field experimental research, as well as considering how economists can better use this research methodology to understand human behaviour and contribute to economic theory (Harrison, 2014, this issue). This paper should be added to the growing list of papers that critique and help shape field experiments going forward.

IV. Moving forward

While there is little doubt that field experiments have deepened our understanding of economic science, pushed forward economic thought, and provided policy-makers with an enhanced tool kit, there is much to be done. We believe that there are four fruitful areas for field experiments. A first key issue, which is also highlighted in List (2011a), is that field experiments should be used in conjunction with economic theory. It has been argued that field experiments, especially in analysing developing world projects and programmes, are too focused on reduced-form estimates and do not tell us much about theory or the structure of the behaviour (Deaton, 2010). A greater emphasis on incorporating field experiments into structural models of human behaviour is a promising

avenue of research (Shearer, 2004; Todd and Wolpin, 2006; Acland and Levy, 2011; Huck *et al.*, 2012; DellaVigna *et al.*, 2012, 2013, 2014).

The second is focused on the partial equilibrium issue of field experiments. Many field experiments focus on the very tight question at hand, but do not have the potential to understand the ripple effects following from the targeted behaviour change. As with any behaviour, there will be an income and substitution effect that needs to be estimated if the total welfare consequences are considered. Take the example of energy efficiency. If we improve the electricity efficiency of a set of households by an intervention, they will then have more wealth to spend on electricity itself (the direct effect), and they will have additional wealth to purchase more energy-intensive products. On the whole, the effects may not be in the same magnitude as many estimates.

The potential for spillover effects puts pressure on economists to link up various data sources. If we intervene to test theory in the field on one behaviour, we should try to understand *ex ante* what the other behaviours are that can be impacted by such an intervention. This full account of the behavioural consequences allows economists to move more towards a general equilibrium model as opposed to solely partial models.

The third issue is sustainability. Many field experiments focus on single interactions or short time horizons as opposed to long-run effects that are potentially important (but, see the work of Landry *et al.* (2010), as a model for moving field experiments from the measurement of short-run effects to a long-run setting). For example, many natural and field experiments aiming to reduce caloric intake cannot disentangle whether caloric information meaningfully changes an individual's consumption or just temporally displaces that consumption (e.g. Wisdom *et al.*, 2010; List and Samek, 2014). The same applies to many behaviours of interest to economists. We also believe that thinking about the strategic interaction of agents and eliciting these in a field-experimental setting is an area of future research.

Allcott and Rogers (2014) offer an interesting experimental discussion of the dynamics of behavioural nudges. The authors use a large-scale natural field experiment to test the effects of social comparison nudges on energy use. They find that home energy reports increase the salience of behavioural nudges to consumers who adjust in the short term by reducing usage. Once the intervention stops, other factors in the consumer's environment steal attention away from the intervention, which means that the conservation gains are not as large as originally anticipated.

The fourth issue is replicability. For instance, Ioannidis (2005) argues that the probability a research claim is true may depend on study power and bias, the number of other studies on the same question, and, importantly, the ratio of true relationships to no relationships in each scientific field. This research calls for more studies to replicate substantive stand-alone findings. For instance, Maniadis *et al.* (2014) found that the prevalence of anchoring is not as large as initial studies had first claimed. The importance of going beyond p-values and presenting the statistical power of the test and the fraction of tested hypotheses that are true associations are key factors to consider when making appropriate inference. Using field experiments to understand the replicability of treatment effects could be a very fruitful area for empirical research more generally.

In sum, field experiments have come a long way in a relatively short period of time. Many policy-makers and practitioners in economics, business, and government are starting to use them to evaluate the causal consequences of a parameter, an intervention, or

a policy. We hope that this issue of the *Review* will help the interested reader understand where field experiments have been used to greatest effect and where the current frontiers are for this method of empirical investigation.

References

- Acland, D., and Levy, M. (2011), 'Habit Formation, Naivet , and Projection Bias in Gym Attendance', mimeo, available at https://gspp.berkeley.edu/assets/uploads/research/pdf/Acland_Levy_working_paper_final.pdf
- Allcott, H., and Rogers, T. (2014), 'The Short-run and Long-run Effects of Behavioral Interventions: Experimental Evidence from Energy Conservation', *American Economic Review*, **104**(10), 3003–37.
- Al-Ubaydli, O., and List, J. A. (2013), 'On the Generalizability of Experimental Results in Economics: With a Response to Camerer', NBER Working Paper No. 19666, National Bureau of Economic Research.
- Andreoni, J. (1990), 'Impure Altruism and Donations to Public Goods: A Theory of Warm-glow Giving', *The Economic Journal*, **100**(401), 464–77.
- Ariely, D., Bracha, A., and Meier, S. (2009), 'Doing Good or Doing Well? Image Motivation and Monetary Incentives in Behaving Prosocially', *American Economic Review*, **99**(1), 544–55.
- Ball, S., Eckel, C., Grossman, P. J., and Zame, W. (2001), 'Status in markets', *Quarterly Journal of Economics*, **116**(1), 161–88.
- Bertrand, M., and Mullainathan, S. (2004), 'Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination', *American Economic Review*, **94**(4), 991–1013.
- Blundell, R., and Costa Dias, M. (2002), 'Alternative approaches to evaluation in empirical microeconomics', *Portuguese Economic Journal*, **1**(2), 91–115.
- Camerer, C. (2011), 'The promise and success of lab-field generalizability in experimental economics: A critical reply to Levitt and List', available at SSRN 1977749.
- Card, D., and Krueger, A. B. (1992), 'Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States', *Journal of Political Economy*, **100**, 1–40.
- Cartwright, N. (2012), 'RCTS, Evidence, and Predicting Policy Effectiveness', in H. Kincaid (ed.), *Oxford Handbook of Philosophy of Social Science*, Oxford, Oxford University Press, 298.
- Deaton, A. (2010), 'Instruments, Randomization, and Learning about Development', *Journal of Economic Literature*, **48**(2), 424–55.
- DellaVigna, S., List, J. A., and Malmendier, U. (2012), 'Testing for Altruism and Social Pressure in Charitable Giving', *Quarterly Journal of Economics*, **127**(1), 1–56.
- Rao, G. (2013), 'The Importance of Being Marginal: Gender Differences in Generosity', NBER Working Paper No. 18748, National Bureau of Economic Research.
- (2014), 'Voting to Tell Others', NBER Working Paper No. 19832, *National Bureau of Economic Research*.
- Dolan, P., and Galizzi, M. M. (2014), 'Getting Policy-makers to Listen to Field Experiments', *Oxford Review of Economic Policy*, **30**(4), 725–52.
- Downs, J. S., Loewenstein, G., and Wisdom, J. (2009), 'Strategies for Promoting Healthier Food Choices', *American Economic Review*, **99**(2), 159–64.
- Engel, C. (2011), 'Dictator Games: A Meta Study', *Experimental Economics*, **14**(4), 583–610.
- Gneezy, U., Leonard, K. L. and List, J. A. (2009), 'Gender Differences in Competition: Evidence from a Matrilineal and a Patriarchal Society', *Econometrica*, **77**(5), 1637–64.
- Goldin, C., and Katz, L. F. (2008), *The Race between Education and Technology*, Cambridge, MA, Belknap Press.
- Graves, P. (2010), *Consumerology: The Market Research Myth, the Truth about Consumers and the Psychology of Shopping*, Boston, MA, and London, Nicholas Brealey Publishing.

- Hallsworth, M. (2014), 'The Use of Field Experiments to Increase Tax Compliance', *Oxford Review of Economic Policy*, **30**(4), 658–79.
- Harbaugh, W. T., Krause, K., and Vesterlund, L. (2002), 'Risk Attitudes of Children and Adults: Choices over Small and Large Probability Gains and Losses', *Experimental Economics*, **5**(1), 53–84.
- Harrison, G. W. (2014), 'Cautionary Notes on the Use of Field Experiments to Address Policy Issues', *Oxford Review of Economic Policy*, **30**(4), 753–63.
- List, J. A. (2004), 'Field Experiments', *Journal of Economic Literature*, **42**(4), 1009–55.
- Heinberg, A., Hung, A. A., Kapteyn, A., Lusardi, A., Samek, A. S., and Yoong, J. (2014), 'Five Steps to Planning Success. Experimental Evidence from US Households', *Oxford Review of Economic Policy*, **30**(4), 697–724.
- Helm, D., and Hepburn, C. (eds) (2009), *The Economics and Politics of Climate Change*, Oxford, Oxford University Press.
- Huberman, B. A., Loch, C. H., and Önçüler, A. (2004), 'Status as a Valued Resource', *Social Psychology Quarterly*, **67**(1), 103–14.
- Huck, S., Rasul, I., and Shephard, A. (2012), 'Comparing Charitable Fundraising Schemes: Evidence from a Field Experiment and a Structural Model', Discussion Paper, Berlin, WZB.
- Hung, A., Parker, A. M., and Yoong, J. (2009), 'Defining and Measuring Financial Literacy', Working Paper WR708, RAND Corporation.
- Ioannidis, J. P. A. (2005), 'Why Most Published Research Findings are False', *PLoS Medicine*, **2**(8), e124.
- Jasper, C. R., and Samek, A. S. (2014), 'Increasing Charitable Giving in the Developed World', *Oxford Review of Economic Policy*, **30**(4), 680–96.
- Landry, C., Lange, A., List, J. A., Price, M. K., and Rupp, N. G. (2010), 'Is a Donor in Hand Better than Two in the Bush? Evidence from a Natural Field Experiment', *American Economic Review*, **100**(3), 958–83.
- Levitt, S. D., and List, J. A. (2007a), 'What Do Laboratory Experiments Measuring Social Preferences Reveal about the Real World?', *Journal of Economic Perspectives*, **21**(2), 153–74.
- — (2007b), 'Viewpoint: On the Generalizability of Lab Behaviour to the Field', *Canadian Journal of Economics/Revue canadienne d'économique*, **40**(2), 347–70.
- — (2009), 'Field Experiments in Economics: The Past, the Present, and the Future', *European Economic Review*, **53**(1), 1–18.
- Neckermann, S. (2014), 'What Field Experiments Have and Have Not Taught Us about Managing Workers', *Oxford Review of Economic Policy*, **30**(4), 639–57.
- List, J. A. (2006), 'Field Experiments: A Bridge between Lab and Naturally Occurring Data', *The B.E. Journal of Economic Analysis and Policy*, **6**(2 - Advances), Article 8.
- (2007), 'On the Interpretation of Giving in Dictator Games', *Journal of Political Economy*, **115**(3), 482–93.
- (2011a), 'Why Economists Should Conduct Field Experiments and 14 Tips for Pulling One Off', *Journal of Economic Perspectives*, **25**(3), 3–15.
- (2011b), 'The Market for Charitable Giving', *Journal of Economic Perspectives*, **25**(2), 157–80.
- Mason, C. F. (2011), 'Are CEOs Expected Utility Maximizers?', *Journal of Econometrics*, **162**(1), 114–23.
- Rasul, I. (2011), 'Field Experiments in Labor Economics', *Handbook of Labor Economics*, **4**, 103–228.
- Samek, A. S. (2014), 'The Behaviorist as Nutritionist: Leveraging Behavioral Economics To Improve Child Food Choice and Consumption', NBER Working Paper No. 20132, National Bureau of Economic Research.
- Maniadis, Z., Tufano, F., and List, J. A. (2014), 'One Swallow Doesn't Make a Summer: New Evidence on Anchoring Effects', *American Economic Review*, **104**(1), 277–90.
- Price, M. K. (2014), 'Using Field Experiments to Address Environmental Externalities and Resource Scarcity: Major Lessons Learned and New Directions for Future Research', *Oxford Review of Economic Policy*, **30**(4), 621–38.

- Reiley, D. H., and List, J. A. (2007), 'Field Experiments in Economics', *The New Palgrave Dictionary of Economics*.
- Rosenbaum, P. R., and Rubin, D. B. (1983), 'The Central Role of the Propensity Score in Observational Studies for Causal Effects', *Biometrika*, **70**(1), 41–55.
- Rosenzweig, M. R., and Wolpin, K. I. (2000), 'Natural "Natural Experiments" in Economics', *Journal of Economic Literature*, **38**, 827–74.
- Sadoff, S. (2014), 'The Role of Experimentation in Education Policy', *Oxford Review of Economic Policy*, **30**(4), 597–620.
- Shearer, B. (2004), 'Piece Rates, Fixed Wages and Incentives: Evidence from a Field Experiment', *Review of Economic Studies*, **71**(2), 513–34.
- Todd, P. E., and Wolpin, K. I. (2006), 'Assessing the Impact of a School Subsidy Program in Mexico: Using a Social Experiment to Validate a Dynamic Behavioral Model of Child Schooling and Fertility', *American Economic Review*, **96**(5), 1384–417.
- Winking, J., and Mizer, N. (2013), 'Natural-field Dictator Game Shows No Altruistic Giving', *Evolution and Human Behavior*, **34**(4), 288–93.
- Wisdom, J., Downs, J. S., and Loewenstein, G. (2010), 'Promoting Healthy Choices: Information versus Convenience', *American Economic Journal: Applied Economics*, **2**(2), 164–78.