3-2020

Experimental approaches in development and poverty alleviation

Yana van der Meulen Rodgers
Rutgers University–New Brunswick

Anthony J. Bebbington
Clark University, abebbington@clarku.edu

Catherine Boone
London School of Economics and Political Science

Jampel Dell'Angelo
Vrije Universiteit Amsterdam

Jean Philippe Platteau
University of Namur

See next page for additional authors

Authors

Yana van der Meulen Rodgers, Anthony J. Bebbington, Catherine Boone, Jampel Dell'Angelo, Jean Philippe Platteau, and Arun Agrawal

WORLD DEVELOPMENT SYMPOSIUM

Experimental Approaches in Development and Poverty Alleviation

Yana Rodgers[1]
Anthony Bebbington[2]
Catherine Boone[3]
Jampel Dell'Angelo[4]
Jean-Philippe Platteau[5]
Arun Agrawal[6]

Pre-Print. Published in *World Development*, March 2020, 127(3), pp. 1-6.
https://doi.org/10.1016/j.worlddev.2019.104807

Abstract

This inaugural *World Development* Symposium on Development and Poverty Alleviation brings together contributions from a range of disciplines, scholars, practitioners, and countries to mark the recognition of Abhijit Banerjee, Esther Duflo and Michael Kremer (BDK) through the 2019 Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel. Contributors examine how BDK's work has changed the methods and study of Development Economics, and their extended impact in other social science and interdisciplinary fields. Although experimental evaluation has had a profound impact on the conduct of much research and policy making, further development of RCT approaches, and collaboration across methods and disciplines, and between scholarship and practice, remain crucial to address the most pressing challenges of sustainability and development.

[1]Department of Labor Studies & Employment Relations, Rutgers University, 94 Rockafeller Road, Piscataway, NJ 08854; T: 848-932-4614; E: yana.rodgers@rutgers.edu
[2]Graduate School of Geography, Clark University, Worcester, MA 01610-1477. E: abebbington@clarku.edu
[3]Department of Comparative Politics, Departments of Government and International Development, Connaught House 604, London School of Economics, Houghton Street, London WC2A 2AE UK. T: 0207 107 5153; E: C.Boone@lse.ac.uk
[4]Institute for Environmental Studies (IVM), Vrije Universiteit Amsterdam. E: jampel.dellangelo@vu.nl
[5]Department of Economics, University of Namur, Belgium. E: jean-philippe.platteau@unamur.be
[6]School for Environment and Sustainability, University of Michigan, Ann Arbor MI 48109. E: arunagra@umich.edu

Introduction

The news that Abhijit Banerjee, Esther Duflo and Michael Kremer (BDK) won the 2019 Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel appeared to us a fitting occasion for *World Development*'s inaugural symposium on Development and Poverty Alleviation. There are at least two reasons: The work BDK have launched is centrally concerned with the mission of the journal. In recognizing their contributions, the Nobel Prize award committee also drew attention to the critical importance of addressing persistent challenges to human and social wellbeing, pressing problems of global poverty, and unequal access to education and health services. Additionally, the methods and approaches championed so ably by BDK have influenced the study of development, particularly its methods and especially among economists. But they have also left their mark on other disciplines and interdisciplinary fields.

BDK's advocacy of experimental methods, and particularly of randomized controlled trials (RCTs), has helped drive home the idea that more rigorous evaluations have substantial potential to question conventional wisdom, generate reliable knowledge, and constitute the foundation for policy improvements to address development challenges more effectively (Bruhn and McKenzie 2009, Banerjee and Duflo 2009, Humphreys and Weinstein 2009). They recognized the potential of experiments for more effective development interventions early, and have worked consistently to refine the approach for implementation in the context of international development and provision of public services (Banerjee et al. 2016, Chupein and Glennerster 2018, Duflo and Banerjee 2017). In using this occasion to publish a symposium devoted entirely to an assessment of experimental approaches, one that is attentive to contributions from scholars and field practitioners, we seek both to recognize BDK's accomplishments and to expand the boundaries of knowledge about experimental methods.

Over the last two decades, randomized controlled trials have gained a "gold standard" reputation in empirical microeconomics and in development economics. Many researchers view them as necessary to establish casual relationships in empirical investigations. A substantial literature relying on experiments has emerged to advance insights based on original field data as well as on systematic reviews and meta-analyses of experimental results in specific domains (Gisselquist and Niño-Zarazúa 2015, Muralidharan 2017).

But influential as randomized controlled trials have been, their criticisms also continue apace (Donovan 2018). Many are related to their small scale, limited temporal extent, external and even internal validity, costs of implementation, ethical oversights, and technocratic orientation. These drawbacks of RCTs are well rehearsed in the literature. In this symposium we have tried to limit rehashing past criticisms. Instead, we have encouraged the contributors to build on or move beyond familiar criticisms in favor of more-forward looking engagements (See also Levitt and List 2009). Indeed, many of the contributions, far from being critiques, instead point to ways in which the experimental approach might be further elaborated: methodologically, institutionally and theoretically.

The contributions to the symposium span a range of voices. Indeed, we made a concerted effort to represent varying orientations in assessments of experimental methods and include newer ideas. Our authors also represent a number of dimensions of diversity: gender, setting

(higher- and lower-income countries), profession (scholars as well as practitioners), seniority, and reputation. Representing a range of backgrounds and views has made for a robust set of articles that are based on authors' own scholarly work and experience and also on reflections about research relying on randomization and field data collection as a scholarly enterprise. As a whole this symposium not only engages with the body of scholarship for which BDK's work has been generative, it speaks more generally also to the necessity and effectiveness (or otherwise) of the experimental approach in alleviating global poverty.

This introduction would be remiss not to highlight that Esther Duflo is only the second woman to have won the Nobel Prize in Economics and the youngest person to do so. Hers is an extraordinary feat given the persistently low representation of women in economics and the chilly climate that women economists continue to face (Lundberg and Stearns 2019). Dr. Duflo is also one of the only Nobel Prize winners to have written specifically about gender equality and women's wellbeing, another key research area integral to the mission of this journal. Her work has contributed both substantively and substantially to the global discourse on the links between economic development and gender equality and the need to reduce the disadvantages faced by women, especially in educational attainment, life expectancy, and labor force participation. In fact, Duflo's (2012) widely-cited review article on women's economic empowerment and economic development – with an emphasis on RCTs and field experiments - concludes that in the face of persistent biases about women's abilities and the lack of a clearly documented relationship between women's empowerment and economic development, active gender-equitable policies continue to be necessary to promote equity in health, education, wages, employment, rights, and political voice.

Contributions in this Symposium

The articles in this symposium engage with a range of issues and arguments related to development and social change. We organize them rather loosely according to the extent to which they respond to central theoretical, methodological, and substantive issues in development and poverty alleviation. Some contributions are more positive about what RCTs have contributed to the study of development. Others remain critical. Some focus on specific sectors. Others are general. Some are concerned with the role of experiments. Others deploy a more distanced analytical lens.

Thus, rather than discussing whether RCTs are an effective means to strengthen knowledge about development and poverty alleviation, five of the contributions in this issue attend instead to the different contexts in which RCTs are embedded. First, Davesh Kapur reflects on a history of different scholarly emphases in the study of development and poverty alleviation over the past several decades. Based on this rapid survey, he suggests that emerging emphases tend to have a half-life of ten to twenty years. He asks whether the current attention to RCTs may run a similar course. Second, Michael Cox's contribution uses theories of human group psychology to consider whether advocates and critics of RCTs – in solidifying their within-group identities – may be missing opportunities for productive scientific interactions and dialog. Moreover, Jean Drèze discusses three important ingredients of sound policy – understanding, value judgements, and deliberation – that go beyond the straightforward formulation of evidence-based policy making.

In some contrast, Luciana de Souza Leão and Gil Eyal adopt a comparative historical perspective and ask why an earlier episode of large-scale randomized trials in development – in the 1960s and 1970s – did not capture the attention of development economists and donors the way the approach advocated by BDK has (See also Leão and Eyal 2019). Their focus echoes some of the analysis on the past and future of experiments offered in Levitt and List who also discuss how the mid-twentieth century experiments, often with the explicit involvement of government agencies, "moved the exploration from plots of land to groups of individuals" (2009: 1). This earlier failure to launch is all the more puzzling given that the contemporary RCTs are characterized by relatively more minor treatments, are of shorter duration, and are often in partnerships with non-governmental organizations rather than governments. Finally, the fifth contribution to explore the broader context of RCTs, by Alexandra Avdeenko and Markus Frölich, examines RCTs in development in relation to experimentation in medicine. It provides an assessment of the key areas where scholars of development interested in RCTs as their method of choice can still learn from counterparts in health. Their contribution recalls some of the points made by Favereau (2016).

Most contributions engage generally with RCTs in terms of their overall strengths and pitfalls. Collectively they bring a range of diverse perspectives in their engagement. Martin Ravallion provides an informative account of the main virtues of RCTs – including their unbiasedness and transparency – as well as their limitations, with the conclusion that a broad range of research methods are crucial for filling remaining knowledge gaps. Christopher Barrett and Michael Carter reflect on the importance of RCTs for development research but caution the reader to consider ethical risks, a high heterogeneity in intervention impacts, and non-classical measurement errors. Some of these concerns can be addressed by better incorporating theorizing into the design of RCT methods. The contribution by David McKenzie provides an elegant response to the question of why care about experiments when it is not RCTs that led to the most prominent success stories of development in our time (China) or historically (UK, US, and Europe). A similar theme animates the contribution by Paul Glewwe who talks of the relationship between theory and experiments in the context of education and when and how they might inform each other.

Seán Muller also acknowledges the positive contributions of RCTs to intellectual inquiry in development economics but questions the contradiction that is posed when assumptions behind causal relationships based on non-experimental research methods are rejected on the one hand, and assumptions for extrapolating research results from experiments are accepted, on the other. Interestingly, his skepticism of the use of machine learning methods to extrapolate experimental results is greeted with more optimism in the contribution by Shawn Cole and co-authors, which views the digital delivery and evaluation of interventions using closed-loop data environments as a promising development in improving the scale and insights of RCTs.

Sarah Baird and co-authors directly address the criticism that RCTs have limited policy relevance. Focusing on the school-based deworming program intervention conducted by Miguel and Kremer (2004), they show how its thoughtful design and attention to the broader policy agenda led to substantive changes both in knowledge and policy outcomes. Their contribution, as also that by Rukmini Banerji and Madhav Chavan on the role of RCTs in enhancing the reading

skills of Indian school children, demonstrate the long term engagement that BDK's experimental approach allows to support policy change and impact.

Closely related, Vijayendra Rao raises concerns about the paradoxical risk that development policies could be driven by the specific demands of scientific technique rather than scientific technique being driven by policy needs. He also points to the limitations that RCTs have in dealing with contextual variation and the fact that one intervention that could work in one region would not necessarily work in a different one.

A number of articles in this symposium address some of the weaknesses of RCTs and discuss alternative approaches. For example, Ingrid Kvangraven critiques the focus of RCTs on individual utility-maximizing behavior and points to the need for using pluralistic methods that account for human agency and socioeconomic context. A similar argument is made in the piece by Sonal Zaveri. Relatedly, the contribution by Janneke Pieters and Stephan Klasen and that of Arjan de Haan and co-authors focus specifically on women's economic empowerment and argue that RCTs have yielded valuable lessons for achieving gender equality, especially regarding microcredit, business services, and education, but more is needed to understand the constraints faced by women, including their unpaid care work and social norms. The thrust of such contributions is less to dismiss RCTs and more to call for productive ways to assess areas of collaborative work that marries the strengths of RCTs to those of alternative research designs and observational data analyses.

In a similar vein, Haroon Akram-Lodhi uses an example from Pakistan to note how rural property relations and asset inequalities can influence and constrain agency in ways that perpetuate poverty and that are not easily captured in RCTs. In not attending to such systemic factors, Fiona Gedeon Achi suggests that RCTs foster future thinking that emphasizes incremental rather than more profound changes.

Another critique of RCTs is internal validity, a concern that Erwin Bulte and co-authors discuss in their piece describing the difficulties of disentangling a treatment effect into the behavioral component and the direct effect of the intervention itself. Using the case of improved seed distribution in Africa, the authors show how observational studies can help to disentangle these components and more clearly identify the intervention effect.

A number of contributions to the symposium focus on how experiments are implemented. Andrew Dillon and co-authors are particularly concerned with data quality and argue that economists have paid less attention to the process of collecting high quality data than they have to randomized variation. Greater investment in research on producing higher quality data would help not only to improve causal inference but would also contribute to poverty reduction. Along these lines, A. Rani Parker and co-authors focus on a large-scale RCT with multistakeholder fora in Uganda that was conducted over three years. They highlight how the technical demands of the experiment and the reality of practitioner worlds forced adjustments around questions of data, measurement of impacts, and continued collaboration.

Pieternella Pieterse's article also cautions that more attention needs to be paid to the implementation of RCTs, especially to design failures and what happens on the ground even

before the data is collected (cf. Muralidharan 2017). Combining qualitative studies with RCTs would help to better understand the local contexts and what happens in the communities where the RCTs take place. A similar argument is made by Naila Kabeer, who points to "RCT misbehavior" and the problems that arise when stakeholders who are involved in RCTs inadvertently influence the outcomes of the experiments according to their own self-interests and constraints. Also exploring some of the methodological concerns associated with RCTs, Aardra Surendran and Awanish Kumar counter the popular view that RCTs have brought empirical work back into economics and that there is a clear binary between theory and empirics, of which RCTs embody the empirical work.

RCTs have received their fair share of criticisms on ethical grounds (Findley et al. 2016). Nimi Hoffmann's essay develops this concern perhaps more than any other in this symposium. Using a dataset of a large number of RCTs, she points to the lack of discussion in published articles of informed consent and some of the challenges in obtaining informed consent in cluster randomized trials involving vulnerable populations. Her arguments make a strong case that researchers conducting RCTs need to be especially attentive to issues of consent given the long history of power asymmetries between researchers in rich countries or in well-endowed research institutions in lower income countries, and the targets of research in experiments. Lennart Kaplan and co-authors also highlight ethical issues, but more from the point of view of research staff located in lower and middle income settings where a large proportion of RCTs have been conducted. Such staff face challenges both to their physical and emotional wellbeing that prevailing political asymmetries may prevent from being fully acknowledged.

Another concern is the limited scope of many RCTs, a question taken up by Sara Stevano. She focuses on the critique that RCTs break up big development problems into small questions and argues that small questions are worth asking if the answers are big. The solution is to bridge the divide between micro-level phenomena and macro-level processes. Macartan Humphreys and Alexandra Scacco also address this issue with a set of suggestions for using evidence from experimental findings for better inferences related to larger theoretical questions. They suggest multi-sited studies that may allow researchers to take contextual variables into account, but concede that to do this, experimentalists must employ theories of similarity and difference across target populations, and strategies for determining which variations constitute impediments to generalization and theory-building.

This conclusion echoes the arguments of other contributors who call for more explicit cross-fertilization and collaboration across the RCT-observational research divide. Rachel Gisselquist argues directly that single experiments should be treated as "single case studies" – thus recognizing the challenges of external validity, generalization, and aggregation from the outset – and treated as hypothesis generating exercises. Aggregating experimental studies across multiple sites which are selected via theoretically-informed processes of case selection, as suggested in comparative methods, can be a way of linking experimental "cases" to theory (see Arceneaux and Nickerson 2009 and Lybbert et al. 2010 for examples). Also addressing the issue of external validity, Martin Williams suggests mechanism mapping as a practical tool for policy makers to integrate experimental evidence from other locations with information about their own local contexts when making choices about scaling up successful interventions (see also Muralidharan and Niehaus 2017). In general, concerns about external validity are widespread in

work on development and are not specific merely to RCTs – they are indeed the subject of substantial scholarship (Acemoglu 2010, Card et al. 2011, Rodrik 2008).

Another methodological issue explored in this symposium, by Sabyasachi Das, is the lack of attention in RCTs to the political process of policymaking, and the need to incorporate issues of political economy into RCTs. C. Austin Davis and Ahmed Mushfiq Mobarak similarly examine the complexities that arise in moving from RCT research results to anti-poverty policy. They attend in particular to the need for tightening this link by changing the practice of RCTs to be more ambitious about treatments, incorporating other rigorous evaluation methods, and scaling up experiments. In the process, they highlight how RCT researchers have sought to address issues of heterogeneity and scaling (see also Subramanian et al. 2018). Philippe Krause and Gonzalo Licona narrate a more successful example of how evidence-based approaches, especially motivated by the success of the RCT-based evaluations of *Progresa*, were mainstreamed in Mexico's policy making and evaluation.

Closer engagement with policy makers and identifying their evidence needs is an example of an innovation to make policy research more inclusive, as argued by Sudhanshu Handa and co-authors in their Transfer Project article. Such engagement is also the concern of Javier Escobal and Carmen Ponce on the basis of their own experiences of conducting RCT-based evaluation within public programs in Peru. They highlight the practical challenges faced by researchers as they develop the relationships of trust with policy makers that are essential for effective RCT design. As they note, this relationship building has implications for how technical capacities for program evaluation are built up within the public sector.

Several authors highlight benefits of RCTs that have been overlooked. In particular, Joana Naritomi and co-authors describe how RCTs create new opportunities for collaboration that do not always happen with other research methods. These collaborations are often cross disciplinary in nature and can also involve work of both academics and policymakers as well as researchers from lower- and higher-income countries, resulting in a greater potential to deepen the knowledge base in development economics. Relatedly, Cyrus Samii highlights benefits of experiments distinct from control of selection bias and argues for how experimentation allows for more deeply engaged learning about policy formulation and implementation.

In contrast, both Kabeer and Akram-Lodhi point to field experiences where RCT-teams have been disinclined to collaborate with qualitative researchers, suggesting that there is still work to be done to foster cultures of research and evaluation that embrace the types of cross-disciplinary approaches central to much work published in *World Development* (while, of course, also recognizing where such collaboration might have its own limits).

Alexandra Hartman and Florian Kern write a strong endorsement of methodological and inferential transparency in best-practice experimental studies (including reporting standards, pre-registration, data sharing, replication, and aggregated evidence), and argue that some of these principles could equally improve the quality and robustness of qualitative research. They view such cross fertilization as an important means to leverage synergies between qualitative and experimental work. Their contribution illustrates this point by proposing a research framework for incorporating some of the strengths of qualitative work (such as interactive development of

data analysis frameworks and attention to scope conditions) into multi-sited experimental research.

Symposium contributors also focus on particular sectors and interventions. The Nobel Prize winners are well known for their RCTs in preventive healthcare in lower-income countries. Jishnu Das provides an informative evaluation of their research on preventive services such as deworming and insecticide-treated bed-nets and how experimental methods have led to new ways of thinking around how to meet the healthcare needs of the poor. Closely related to health is nutrition, an area addressed by Purnima Menon and co-authors. They discuss a range of evaluation methods, including experimental approaches, to help inform a global nutrition initiative that has provided benefits to over 16 million women and children.

Referring again to health and nutrition, Agnes Quisumbing and co-authors discuss how their randomized trials have addressed three common criticisms of RCTs: their black box nature, external validity, and scalability. Addressing these criticisms, they suggest, goes hand in hand with delivering well-designed and effective health and nutrition programs.

The contribution by Alain de Janvry and Elisabeth Sadoulet focuses directly on agriculture and documents how RCTs in agriculture have transitioned from the lab to the field. They highlight especially the impediments that RCTs in agriculture must address to perform effectively in poor countries. Their analysis recalls the poignant title of a recent paper that asks if field experiments can "return agricultural economics to the glory days" (Herberich et al. 2009: 1259).

In the area of education, Rafael de Hoyos provides a positive assessment of experimental evaluations that have been used to design incentives to improve student learning outcomes. He also suggests innovations to complement education systems in order to reach lagging students. More broadly, peer effects in education and health that have been documented by the Nobel prize winners are also applicable in the area of agriculture. In this symposium, Alan de Brauw and Vivian Hoffmann explore how peer effects can be used to more effectively disseminate information on new seeds and improved technologies to smallholders.

Although development economists have adopted RCTs widely, their use remains limited in scholarship on governance of conservation, environment, and sustainability (but see List and Price 2016, Grillos et al. 2019, Pynegar et al. 2018, Voors et al. 2011). Several contributors to the symposium discuss both the need and strategies for RCTs in conservation, climate change, and environmental policy. As discussed by Nigel Asquith based on his experience in Bolivia, it is indeed possible to implement a successful large-scale RCT for an incentive-based conservation program if the implementers are able to replicate a proven conservation intervention at scale, effectively evaluate the outcomes, put their reputations on the line, and have patience. Francisco Alpízar and Paul Ferraro articulate a compelling argument for the use of RCTs to bridge the knowledge gap on how poverty programs affect the environment and conversely, how environmental interventions affect poverty. In their case study of the UN's Green Climate Fund, Jyotsna Puri and co-authors examine how rigorous impact assessments and RCT-based findings can be used to build the evidence base on the impact of climate change interventions in multilateral agency settings.

Of course, it is not just in relation to conservation and environmental policy that the use of RCTs is not as widespread as it has become in relation to policies for provisioning in health, education, and social assistance. John Quattrochi and co-authors examine why experiments have found limited uptake in humanitarian settings, and illustrate how some of the obstacles can be addressed, drawing upon their experience with large sale humanitarian programs in the Democratic Republic of Congo.

In another positive assessment of the value of RCTs in better understanding how to mitigate climate change, Stefano Carattini and co-authors point to how RCTs have the potential to deepen researchers' understanding of the forces that drive the diffusion of non-normative behaviors and technologies aimed at lowering greenhouse gas emissions. Much of the reviewed research, based on social network analysis and field experiments in developing countries, has important lessons for policy makers across the development spectrum (see also Breza 2016). Saleem Ali, however, is more circumspect regarding the place of RCTs in orienting environmental governance and planning, and dealing with urgent ecological challenges. He worries that they foster more technocratic approaches to defining policy, squeezing out space for participation and adaptive planning, and also that the time required for RCTs can delay responses to environmental problems that are more urgent in nature.

Microcredit is another subject of close scrutiny in RCTs (Banerjee et al. 2015). Jonathan Morduch argues that RCTs are helpful in evaluating the benefits of microcredit in alleviating poverty, but concerns about their narrow focus and external validity limit their usefulness. Similar to other authors in this symposium, he points to the importance of a plurality of methods in assessments of the effectiveness of microcredit.

Finally, Heidi McGowan provides some useful advice to those seeking to submit RCT-based research papers for publication. Her contribution, which offers suggestions not only for writing up the final results but also for planning and implementing experimental approaches in development, will be useful to any author seeking insights into how the editorial process works for *World Development* and comparable journals.

Conclusion

Overall, the symposium contributions render insight into two major puzzles associated with RCTs. First, because of their focused and carefully designed/chosen experimental settings and parameters, RCTs enable greater precision in estimating causal effects within a sample population than other methods with comparable sample sizes but without randomization. However, because experimental work often does not explicitly take into account how contextual conditions of experiments may be affecting treatment effects, it is difficult to know how outcomes of an RCT in a given context will change in other contexts, especially in future time periods where RCT replications may encounter unpredictable conditions. Findings of a given RCT, therefore, may be difficult to generalize beyond a given experimental setting. Second, estimating causal effects with experimental methods may not be a necessary condition for better development policy. For many development policies, consistency in how a given set of variables

predicts changes in outcomes of interest may also be sufficient to guide policy and interventions under many circumstances.

In thinking about BDK's work and how it has contributed to the rise of randomized controlled trials and experimental methods in shaping development interventions and influencing the study of development, it is worth returning to the transformations in the structural context of experiments and the contexts in which experiments have occurred. These changes include, among others, the decline of the Washington Consensus, the fragmentation of the aid world, the rise of philanthropists interested in outcomes of interventions they sought to support, and the promise of objectivity embodied in the small-scale and technocratic orientation of current day experiments as Leão and Eyal (this issue) highlight. These changes in the social and political context of development interventions have clearly played a supporting role in helping experiments in social policy and development find the appeal and popularity they have discovered – regardless of whether economists and development scholars are best described as theorists, engineers, craftsmen, or plumbers (Duflo 2017:2).

The obvious question, nonetheless, remains how this power to "make big differences" (Banerjee and Duflo 2009: 174) can be used responsibly to support greater wellbeing – not just of people and societies, but also of the environment and ecosystems. Development and sustainability are today intertwined as never before in history. Challenges of climate change, emissions reduction, biodiversity decline, and ecosystem loss and erosion occur at scales that are only partially amenable to the requirements of randomization in experimentation. Addressing these challenges requires the kind of methodological pluralism and triangulation across different strategies of knowing for which many of the most insistent contributors to this symposium have called. It also requires a willingness to look at the patterns, relationships, and socio-environmental outcomes of both social and environmental interventions and policies, as Alpízar and Ferraro highlight in their contribution.

The high stakes of decisions in policy making around development and poverty alleviation, the uncertainties and non-linearities inherent in the dynamics of coupled human and natural systems, and the unpredictable range of outcomes that external interventions catalyze, when coupled with value differences and interest conflicts, almost beg for collaboration among those who view RCTs as promising more rigorous scientific knowledge for solving real world problems, and others who see solutions to problems as being inescapably about engagement, deliberation, and consultation rather than technical expertise. Obviously, there is no necessary conflict between conducting randomized trials and consulting and engaging those whose wellbeing such trials seek to enhance. Indeed, much research by BDK is emblematic of rigorous knowledge generation as well as deep, long-term, place-based engagement.

This value of collaboration among different types of scholarship, as well as between scholars and practitioners, is expected to become increasingly important as contemporary RCTs are scaled up – both in terms of their spatio-temporal coverage and their consequential treatments. Scaling up experiments will likely encounter similar political-economic interests invested in preserving the socio-economic and environmental status quo that challenged the first wave of RCTs in development. This value of collaboration, across methodological divisions, also can help overcome the parochial advocacy of RCTs or other methods that sees in its preferred candidates the only way to generate reliable knowledge – whether to advance theory or to guide

development interventions. Finally, this value of collaboration, spanning disciplinary boundaries, is the foundational ground on which to build and implement solutions to sustainability problems for which time is running out.

References

Acemoglu, D. (2010). Theory, general equilibrium, and political economy in development economics. *Journal of Economic Perspectives*, *24*(3), 17-32.

Arceneaux, K., & Nickerson, D. W. (2009). Who is mobilized to vote? A re-analysis of 11 field experiments. *American Journal of Political Science*, *53*(1), 1-16.

Banerjee, A.V. and Duflo, E., 2009. The experimental approach to development economics. *Annual Review of Economics*, *1*(1), pp.151-178.

Banerjee, A. V., Duflo, E., & Kremer, M. (2016). The influence of randomized controlled trials on development economics research and on development policy. In *The State of Economics, The State of the World Conference at the World Bank*. Washington DC: The World Bank.

Banerjee, Abhijit, Dean Karlan, and Jonathan Zinman. 2015. Six randomized evaluations of microcredit: Introduction and further steps. *American Economic Journal: Applied Economics* 7 (1): 1–21.

Breza, E. (2016). Field experiments, social networks, and development. In *The Oxford handbook of the economics of networks*. Oxford University Press.

Bruhn, M., & McKenzie, D. (2009). In pursuit of balance: Randomization in practice in development field experiments. *American economic journal: applied economics*, *1*(4), 200-232.

Card, D., DellaVigna, S., & Malmendier, U. (2011). The role of theory in field experiments. *Journal of Economic Perspectives*, *25*(3), 39-62.

Chupein, T., & Glennerster, R. (2018). Evidence-informed policy from an international perspective. *The ANNALS of the American Academy of Political and Social Science*, *678*(1), 62-70.

Donovan, K. P. (2018). The rise of the randomistas: on the experimental turn in international aid. *Economy and Society*, *47*(1), 27-58.

Duflo, E. (2017). Richard T. Ely Lecture: The Economist as Plumber. *American Economic Review*, *107*(5), 1-26.

Duflo, E. 2012. "Women empowerment and economic development." *Journal of Economic Literature* 50 (4): 1051-79.

Duflo, E., & Banerjee, A. (Eds.). (2017). *Handbook of Field Experiments* (Vol. 1). Amsterdam: North Holland.

Favereau, J. (2016). On the analogy between field experiments in economics and clinical trials in medicine. *Journal of Economic Methodology*, *23*(2), 203-222.

Findley, M., Nielson, D., & Desposato, S. (2016). Obligated to Deceive? Aliases, Confederates, and the Common Rule in International Field Experiments. *Ethics and Experiments. Problems and Solutions for Social Scientists and Policy Professionals*, 151-70.

Gisselquist, R. M., & Niño-Zarazúa, M. (2015). What Can Experiments Tell Us About How to Improve Government Performance?. *Journal of Globalization and Development*, *6*(1), 1-45.

Grillos, T., Bottazzi, P., Crespo, D., Asquith, N., & Jones, J. P. (2019). In-kind conservation payments crowd in environmental values and increase support for government intervention: A randomized trial in Bolivia. *Ecological Economics*, *166*, 106404.

Herberich, D. H., Levitt, S. D., & List, J. A. (2009). Can field experiments return agricultural economics to the glory days?. *American Journal of Agricultural Economics*, *91*(5), 1259-1265.

Humphreys, M., & Weinstein, J. M. (2009). Field experiments and the political economy of development. *Annual Review of Political Science*, *12*, 367-378.

de Souza Leão, L., & Eyal, G. (2019). The rise of randomized controlled trials (RCTs) in international development in historical perspective. *Theory and Society*, *48*(3), 383-418.

Levitt, S. D., & List, J. A. (2009). Field experiments in economics: The past, the present, and the future. *European Economic Review*, *53*(1), 1-18.

List, J. A., & Price, M. K. (2016). The use of field experiments in environmental and resource economics. *Review of Environmental Economics and Policy*, *10*(2), 206-225.

Lundberg, S. and J. Stearns. 2019. “Women in economics: Stalled progress." *Journal of Economic Perspectives* 33 (1): 3-22.

Lybbert, T. J., Galarza, F. B., McPeak, J., Barrett, C. B., Boucher, S. R., Carter, M. R., ... & Mude, A. (2010). Dynamic field experiments in development economics: Risk valuation in Morocco, Kenya, and Peru. *Agricultural and Resource Economics Review*, *39*(2), 176-192.

Miguel, E. and M. Kremer. 2004. Worms: identifying impacts on education and health in the presence of treatment externalities. *Econometrica* 72 (1): 159-217.

Muralidharan, K. (2017). Field experiments in education in developing countries. In *Handbook of economic field experiments* (Vol. 2, pp. 323-385). North-Holland.

Muralidharan, K., & Niehaus, P. (2017). Experimentation at scale. *Journal of Economic Perspectives*, *31*(4), 103-24.

Pynegar, E. L., Jones, J. P., Gibbons, J. M., & Asquith, N. M. (2018). The effectiveness of Payments for Ecosystem Services at delivering improvements in water quality: lessons for experiments at the landscape scale. *PeerJ*, *6*, e5753.

Rodrik, D. (2008). The new development economics: we shall experiment, but how shall we learn?. Faculty Research Working Papers Series RWP08-055, John F. Kennedy School of Government, Harvard University. Cambridge, MA.

Subramanian, S. V., Kim, R., & Christakis, N. A. (2018). The" average" treatment effect: A construct ripe for retirement. A commentary on Deaton and Cartwright. *Social Science & Medicine (1982)*, *210*, 77.

Voors, M., Bulte, E., Kontoleon, A., List, J. A., & Turley, T. (2011). Using artefactual field experiments to learn about the incentives for sustainable forest use in developing economies. *American Economic Review*, *101*(3), 329-33.