Experimental Evidence, Scaling and Public Policy: A Perspective from Developing Countries

ANANDI MANI[1]

I highlight two important factors particular to less-developed countries (LDCs) that can bias evidence generation and contribute to the 'voltage drop' in program benefits, moving from field research experiments to policy implementation at scale. The first is the non-linear increase in *information processing* and coordination costs associated with upscaling in LDCs, given *limited state capacity* and rigid organizational hierarchies. The second is *political bias* in the choice of programs considered for rigorous evaluation itself, resulting in distorted evidence and policy choice. These two factors raise considerations that complement the economics-based approach outlined by Al-Ubaydli et al's (this issue) in the quest for more rigorous, evidence-based policy.

[1] Blavatnik School of Government, University of Oxford [Email: anandi.mani@bsg.ox.ac.uk]. I would like to acknowledge several people whose ideas have influenced my thinking on the issues discussed in this article over the years: Sayantan Ghosal, Lakshmi Iyer, Karla Hoff, Sendhil Mullainathan, Dilip Mookherjee and Debraj Ray. The author is also grateful to Sharun Mukand for many helpful conversations in the course of writing this piece.

Over half a century ago, Hirschmann (1968) observed that solutions to underdevelopment were likely to be particularly challenging. In particular, he argued that given that the problem of conquering underdevelopment was likely "multifaceted, tangled, and deep-rooted", and hence suggested a different approach

> "*What if the fortress of underdevelopment, just because it is so formidable, cannot be conquered by frontal assault? In that unfortunately quite common case, we need to know much more about ways in which the fortress can be surrounded, weakened by infiltration or subversion, and eventually taken by similar indirect tactics and processes. And I suggest that the major contribution to our knowledge of economic development must now come from detailed studies of such processes*."

Half a century later, the above passage sounds prescient. Over the last two decades, academics have made conscious efforts to address the challenges of poverty and under-development through micro-level studies and experimentation. As the scientific background paper for the 2019 Nobel prize makes clear "in just two decades, the pioneering work by this year's Laureates has turned development economics — the field that studies what causes global poverty and how best to combat it — into a blossoming, largely experimental field."

Policy interventions to reduce poverty based on such experimental studies have had some success. But it has been a mixed success at that, and general lessons remain elusive. Among other things, policies that seem to work initially, often end up disappointing when scaled up (Muralidharan and Niehaus, 2017). These hard realities makes Al-Ubaydli et al's (this issue) contribution especially relevant. Given all the attention with the recent Nobel prize, it is particularly important that issues of upscaling based on an experimental approach are critically analysed so as to fully realize its potential to support evidence-based policymaking.

In this paper, I will focus on addressing some of the issues raised by that paper from the perspective of developing countries. In what follows, I highlight what I see as two important considerations in discussions about scaling in such contexts. The first is the implications of *limited state capacity* for tackling the scaling-up challenges highlighted by the authors. The second is the wider issue of the *politics of project selection* and how that may distort the process of evidence-based policy making through experimentation.

A. Limited State Capacity and Scaling in Developing Countries

According to Al-Ubaydli et al's paper, an important reason why the move from the field experiment to actual policy implementation results in a large "voltage drop", is due to problems with the delivery of the treatment. In particular, they argue that "if

the original study involved a certain delivery, in a new situation when the program is delivered incorrectly one cannot expect the measured treatment effect to mirror that measured effect at smaller scale." This is a useful way to characterize the problem of weak delivery with scaling up of the experiment.

However, being aware of constraints on program delivery does not mean that these can be overcome simply by being more vigilant. Two kinds of constraints interact with each other to make the 'voltage drops' especially likely in a developing country context. The first is the information processing challenge that is ubiquitous to all organisations. When projects are upscaled, the information processing challenge makes the cost of monitoring and coordinating increase in a non-linear way. This disproportionate increase in information processing costs cannot be solved by altering the underlying organisational architecture alone. This technical inefficiency associated with any scaling effort, whether in a developed or developing country context, is compounded by a second constraint unique to the latter – namely the challenge of poor state capacity. Below I elaborate on both of these challenges.

(i) Scaling, Bounded Rationality and the Architecture of Governance

There is a natural informational constraint that sets an upper bound on the extent to which scaling can take place without any voltage drop - *even if* program delivery, dosage and the level of monitoring remain unchanged. This is related to the higher cost of processing information that policymakers face, in scaling up.

Such information processing costs rise for two sets of reasons. First, bounded rationality in the ability to process information implies that the time taken to make a decision rises when there is more information to process, perhaps non-linearly so. When scaled up, all projects implemented by government have to move up through multiple layers of bureaucratic hierarchy and decision making, which increases the time taken to arrive at any decision. Second, this transmission through multiple layers will also increase communication costs. These costs include not just the direct costs of communicating, but also the costs associated with errors in transmission of information across various levels of a hierarchy. Such costs can be affected directly by the sequencing of decision making within government.

To take a simple example, consider a scaling up of an intervention from 10 project locations to a hundred project locations. Each project location will incur the same fixed cost of transmitting information up one level of the decision-making hierarchy. This higher fixed cost of information transmission up the government hierarchy scales up quite easily. We may think that there should be a way in which government efficiency can easily be scaled up, by mimicking the governance structure (e.g. single researcher evaluating 10 projects) that existed prior to scaling up, by hiring 10 decision making

policymakers evaluating the 100 projects. An early argument about why this would not be easy was made by Kaldor (1934, p.68)

> "You cannot increase the supply of coordinating ability…as it is the essence of co-ordination that every single decision should be made on a comparison with all the other decisions made or likely to be made; *it must therefore pass through a single brain* (italics mine)."

This intuition was formalised in a series of classic papers by Radner (1992) and Radner and Van Zandt (1997), who showed how the costs of information processing and coordinating actions result in diminishing returns. Thus, given bounded ability to process information, scaling up will result in a non-linear increase in costs (see Van Zandt, 1999 for a survey).

Further, decision makers in developing countries are likely to be low-level public sector functionaries who are poorly paid, live pay cheque to pay cheque and are preoccupied with financial and other worries. Scaling of any experiment as it becomes widely adopted is likely to exacerbate these information processing costs because of some of these "scarcity" related issues faced by poorly paid bureaucratic functionaries (see Mani et al, 2020 for a discussion).

Some of the scaling costs described above could be minimized by altering the architecture of governance (see Sah and Stiglitz, 1986). However, bureaucracies that are often responsible for implementing any upscaling are notoriously rigid, if not downright sclerotic. This is especially true in developing countries because there is a worry that allowing flexibility in organizational design will make the bureaucracy more vulnerable to corruption. Such vulnerability is directly linked to the state's capacity to monitor behaviour through enforcement of checks and balances – which brings me to my second point.

(ii) State Capacity and Development

Governance is a function of the effectiveness of the bureaucracy as well as the direction set by the political leadership. In discussions on scaling, this underlying state capacity to ensure that public officials implement any policy recommendations is taken as a given. This may well be an appropriate assumption when discussing scaling of policy in developed countries – but much less so in the context of the developing world. Government agencies are poorly resourced, lack direction and there is a large heterogeneity in the quality and motivation of public sector workers. To take one common example, consider teachers. They are public sector employees in many countries, both developing and developed. However, while teacher absenteeism is rife in developing countries, it is not an issue in developed countries (see Duflo, Hanna and Ryan, 2012). These differences in effectiveness of public sector employees can

handicap governments delivery of any proposed "treatment" in the schooling curriculum in a developing country, unlike in a developed country.

At first glance it may seem that the problems of poor state capacity would exist even prior to scaling up. If so, it would seem that there is no reason why these effects would worsen with scaling. However, that need not be the case when there is heterogeneity in the quality of public officials implementing the policy. Some of these bureaucrats may be genuinely motivated and require no extra supervision, while others will. Given that the researcher is likely to be poorly informed about the extent of heterogeneity in bureaucratic quality, scaling up will result in a voltage drop, the size of which could be difficult to predict or prevent.

Moreover, when experiments are scaled up in the absence of state capacity, it is no longer possible for a (boundedly rational) supervisor to monitor these less motivated bureaucrats. As a result of which, as the experiment is scaled up, it will not be able to draw on additional high-quality supervisors. Ipso facto, the negative effects of weak state capacity can no longer be masked and the treatment effects will weaken as the experiment is scaled up.

Al-Ubaydli et al (this issue) provide us with a long checklist (around 40 points) and 12 proposals to ensure that scaling does not result in a voltage drop. While this checklist is useful, it would perhaps be even more so for a developing country context if it were shortened to include the most essential points – especially given poorly paid and poorly trained bureaucratic functionaries. If not, the worry is that a poorly motivated bureaucracy with low state capacity will shrug its hands and not take any element of the checklist seriously (Gawande, 2010).

B. Experimental Evidence and the Political Economy of Project Selection

In recent times, the term 'evidence-based policy' has become something of a buzz-word in policy circles. This term assumes a systematic and scientific approach to policy choice, but in fact there is an important bias that is overlooked when we make such an assumption – a bias in the menu of policies for which evidence is solicited by policy makers in the first place. This bias arises because garnering evidence about the usefulness of a policy often has *politica*l implications.

Governments are forward-looking enough to try and anticipate the likely political consequences of adopting any particular policy. We should not expect that they will risk embarking on a path that may upset the political equilibrium against their favour. For instance, even if evidence shows that machine learning algorithms can do better than judges in predicting which under-trial criminals can safely be let out on probation (Kleinberg et al, 2018)), governments could be very reluctant to adopt such an approach given the political risks associated with mistakes in prediction. (The

resistance of judges themselves to such a change could be an equally important impediment to adoption too, of course).

Often enough, governments and the politicians who head them also have strong ideological preferences for particular policy instruments and outcomes. Such strong preferences may also come from public sector employees or party vote banks themselves.[2] Implementing certain policy experiments could generate evidence in support of a policy that the government does not agree with. Such evidence may help form a lobby that supports the scaling up of the policy and the government may find itself backed into a corner, forced to upscale implementation. Anticipating such possibilities, governments would strategically reject giving regulatory approval for the upscaling for evidence generation.[3]

As Berk et al (1985, pp. 396) have argued in discussing the record of experiments with social policy in the United States:
"The political process too often settles the question by ideological fiat. Can an old-fashioned conservative dare to find out whether deregulation really leads to greater efficiency? Can an old-fashioned liberal dare to find out whether unions improve the workers lot? For these reasons rigorous experiments in social policy are rarely embraced (let alone funded) by public officials."

Given all of the above considerations, governments will go out of their way to try and collect information only in ways that do not politically constrain them in the future. In my view, there is insufficient recognition, perhaps even knowingly so, of the bias this can create in the body of evidence generated. Our focus on scaling and evidence-based evaluation often obscures and distracts our attention from this important issue. As a result, there is a danger that especially when it comes to policy advice in the political arena, the most important decisions are made *prior* to the scaling stage or experimentation itself.

A second reason may have to do with the soft budget constraint and the nature of public organisations (Kornai, Maskin and Roland, 2003).[4] A policymaker may fear that by incurring the initial cost of experimentation with a particular experiment may make it unviable for a government to not proceed with scaling.

---

[2] To take one concrete example of the latter, public school teachers in the United States are unionized and prefer higher pay and smaller class sizes. They are also big contributors and supporters of the Democratic Party. On several occasions, unions have fiercely resisted school districts from experimenting with school choice and charter schools (Jason, 2017 and Salisbury, 2005).

[3] The game theoretic arguments for how this may occur in the broader context of policymaking were first described by Coate and Morris (1999) and later elaborated by Majumdar and Mukand (2004).

[4] The literature on soft budget constraints was developed in the context of cost overruns amongst public sector organisations (often in socialist economies) and was formally addressed in a classic contribution by Dewatripont and Maskin, (1995).

This problem of political selection of experiments through the regulatory approval process is a particularly important concern in the field of development. This is because there is often a symbiotic relationship between researchers and the government in the implementation of the experiment. The researcher's ability to conduct experiments at scale often requires cooperation from host country governments. It is not just that governments in question have to give the regulatory go-ahead for such research; experiments at scale also often need the active use of government facilities and resources. For instance, schooling experiments reported and conducted by Muralidharan and Sundararaman (2015) required not just permission from the Andhra Pradesh government, but also for the government to be get its school administrators, teachers and the local government authorities consenting to cooperate in the intervention. In other instances, we often have the host government directly supplying resources to conduct experiments (e.g. IGC collaborations with state government of Bihar in India). These symbiotic relationships between the researcher and the policy making arm of government are often necessary to generate good policies. However, it would be a mistake to think that these are innocuous and do not skew the kind of policies for which evidence is generated at scale.

A more direct reason for politically-motivated selection is that governments do not like being evaluated themselves, for fear of being judged as falling short. Of course, this is true everywhere, but more so in developing countries, given fewer automatic checks and balances. In contrast, developed countries commonly have both more institutional constraints to government power, and a strong media that acts as a watchdog. Neither of these constraints operate with similar effectiveness in developing countries. Furthermore, researchers themselves may choose not to study issues that are politically 'sensitive' or against the ideological predilection of the government, for fear of jeopardizing potential future partnerships.[5]

Governments always can (and will) allocate resources towards the study of those projects that they would like to undertake. However, what we have identified here is a particularly subversive way in which governments can do this – by affecting the *menu* of policies on offer. What is potentially worrying is that the process of evidence-based evaluation (and discussions of scaling up) has the unintended side-effect of providing intellectual cover for a government's politically motivated policy choices. Political opposition to the menu of policies on offer may be dampened as a result of this, because citizens (and researchers) become caught up with evaluating the policies on the menu, rather than questioning what is on the menu itself.

[5] For a recent example of such a politically sensitive policy in the context of Rwanda examined using an experimental approach, see Blouin and Mukand(2019).

The above point may be at least a partial explanation for a question that has been raised by some academics on the kinds of issues that development researchers using the experimental approach have focused on. Let me elaborate. The majority of field experiments (as well as scaling) in developing countries have dealt with issues such as the importance of conditional cash transfers, improved stoves for cooking, treated bed nets, school meals, HIV immunization, safe water storage, women's empowerment and so on (see Table 1, Vivalt(2019) for a comprehensive list). Pritchett (2016) has questioned this focus. In his view, these issues do not seem to be crucial for reducing poverty and catalysing development, given the experience of recent developmental success stories in South Korea, Taiwan or China. He then questions why it is that a generation of the best and brightest scholars of development, have spent time evaluating and studying such interventions.

Our preceding discussion suggests a possible answer. Field experiments and rigorous impact evaluation of some of the most important interventions for development is not simply because they are technically infeasible, but perhaps because governments do not offer them on the menu for researchers to evaluate them. While Al-Ubaydli et al's framework is a useful step forward in that it analyses the problem "through the lens of economic incentives and markets", it misses the role of political incentives in the decision to permit introduction and scaling up of experiments. These factors are likely to be especially important in the context of developing countries where government policymaking is likely to be more influenced by political imperatives than in developed countries.

C. Conclusion

Experiments are ubiquitous in research in development economics. Many of the points raised by Al-Ubaydli et al are useful and deserve to be incorporated as standard practice. However, in this review my aim was to draw on the lessons from developing countries to critically discuss some of the issues that need to be addressed more systematically.

I raised a couple of issues that I believe the literature on field experiments for evidence-based policymaking neglects. The first is the issue of state capacity when discussing issues related to scaling. In the developing country context where information processing is costly and bureaucratic functionaries find it difficult to process information, scaling may result in a large voltage drop in the size of the treatment. A second and equally important challenge for evidence-based policy is politically-motivated selection in the menu of policies that evidence is sought for. There is often the presumption that allowing for evidence-based policymaking will result in the best policies being chosen. No doubt, experimentation-based learning is valuable and may be even necessary for the design and implementation of good policy in the absence of local knowledge. However, the process of knowledge production

itself goes through a political selection filter that may bias evidence generation and learning on which policies can speed up the process of development, poverty reduction and growth.

References:

Omar Al-Ubaydli, Min Sok Lee, John List, Claire Mackevicius and Dana Suskind, 2020, "How Can Experiments Play a Greater Role in Public Policy? 12 Proposals from an Economic Model of Scaling", Working Paper

Berk, R.A., Boruch, R.F., Chambers, D.L., Rossi, P.H. and Witte, A.D., 1985. Social policy experimentation: A position paper. *Evaluation Review*, *9*(4), pp.387-429.

Besley, T. and Ghatak, M., 2005. Competition and incentives with motivated agents. *American economic review*, *95*(3), pp.616-636.

Blouin, A. and Mukand, S.W., 2019. Erasing ethnicity? Propaganda, nation building, and identity in Rwanda. *Journal of Political Economy*, *127*(3), pp.1008-1062.

Coate, S. and Morris, S., 1999. Policy persistence. *American Economic Review*, *89*(5), pp.1327-1336.

Duflo, E., Hanna, R. and Ryan, S.P., 2012. Incentives work: Getting teachers to come to school. *American Economic Review*, *102*(4), pp.1241-78.

Committee for the Prize in Economic Sciences in Memory of Alfred Nobel, 2019. Understanding Development and Poverty Alleviation. https://www.nobelprize.org/uploads/2019/10/advanced-economicsciencesprize2019.pdf

Gawande, Atul, 2010, *The Checklist Manifesto: How to Get Things Right*, Profile Books, London.

Hirschman, A.O., 2014. *Development projects observed*. Brookings Institution Press.

Hirschman, A.O., 1968. "Foreword," pp. vii–viii, in Judith Tendler, *Electric Power in Brazil: Entrepreneurship in the Public Sector*, Harvard University Press, pp. vii-xii.

Kaldor, N., 1934. The equilibrium of the firm. *The Economic Journal*, *44*(173), pp.60-76.

Kleinberg, J., Lakkaraju, H., Leskovec, J., Ludwig, J. and Mullainathan, S., 2018. Human decisions and machine predictions. *The quarterly journal of economics*, *133*(1), pp.237-293.

Kornai, J., Maskin, E. and Roland, G., 2003. Understanding the soft budget constraint. *Journal of economic literature*, *41*(4), pp.1095-1136.

Majumdar, Sumon and Mukand, Sharun(2004), Policy Gambles, *American Economic Review 94*(4), pp.1207-1222.

Mani, A. and Mukand, S., 2007. Democracy, visibility and public good provision. *Journal of Development Economics*, *83*(2), pp.506-529.

Mani, A., Mullainathan, S., Shafir, E. and Zhao, J., (forthcoming, 2020). Scarcity and cognitive function around payday: A conceptual and empirical analysis. *Journal of Association of Consumer Research*.

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

Muralidharan, K. and Sundararaman, V., 2015. The aggregate effect of school choice: Evidence from a two-stage experiment in India. *The Quarterly Journal of Economics*, *130*(3), pp.1011-1066.

Pritchett, Lant, 2014. Is Your Impact Evaluation Asking Questions that Matter: A Four Part Smell Test?, Center for Global Development: Commentary and Analysis.

Radner, R., 1993. The organization of decentralized information processing. *Econometrica*, pp.1109-1146.

Sah, R.K. and Stiglitz, J.E., 1986. The architecture of economic systems: Hierarchies and polyarchies. *The American Economic Review*, pp.716-727.

Salisbury, David (2005), "School Choice: Learning from other Countries", Commentary, Cato Institute

Van Zandt, T., 1999. Decentralized information processing in the theory of organizations. In *Contemporary economic issues*(pp. 125-160). Palgrave Macmillan, London.

Vivalt, E., 2020. "How Much Can We Generalize from Impact Evaluations?". *Journal of the European Economics Association*.

Jason, Zachary (2017), “The Battle over Charter Schools”, Harvard Ed Magazine, Harvard Graduate School of Education, Summer 2017