

Social Learning and the World Bank

Tom Kenyon & Alberto Criscuolo

Prepared for the First Hirschman-Colorni Conference
Boston University, 6-7th October 2017

Introduction:

How should we think about Hirschman's legacy for development practice? Those of us who earn our livings designing, supervising and occasionally writing about development projects often feel ourselves subject to the pull of two conflicting intellectual tendencies. On the one hand, we are faced with clear evidence that, in important respects, things are getting better – especially with respect to basic living conditions.¹ On the other, we are aware of a consensus that feasible development options vary across countries and over time and that the confidence which existed two decades ago around the conditions required for economic growth has dissipated. The challenge then, as a leading practitioner puts it, is to 'find an orienting framework that can fill the gap between hubris on the one hand and despair disguised as humility on the other.'² What is required is less a set of prescriptions than support for a process of structured, cumulative experimentation that admits the existence of uncertainty but at the same time acknowledges and builds on existing knowledge.³ In this context, Hirschmanian themes of social learning continue to be of relevance.

Defining the problem:

Achieving structural change in the long-run growth of an economy requires that a wide range of market failures and government failures be addressed, over time, across multiple sectors. Human capital must be built; infrastructure must be designed and constructed; access to finance must be enabled; technology must be transferred. A broad base of firms must come into existence; must discover true costs of production; must acquire capabilities; and must move toward the global quality frontier. Policy instruments that support firms in accomplishing these tasks must also discipline them, so that the most productive firms expand. Similar considerations apply to many other forms of public goods provision.

¹ See inter alia, Charles Kenny, Getting Better.

² Levy p. 8.

³ 'The case for agnosticism... has three steps: that feasible options vary across countries and over time; that over the short to medium run the resulting choices can widen divergence in patterns of governance across countries; but that, over the long run, the power of cumulative causation can lead to convergence.' (Levy p. 210)

This is a daunting challenge. It is also a dynamic process. In an ideal world, policy-makers would evaluate the social returns on public investments in physical, human and knowledge capital and calibrate supply to the point at which returns on each are equalized. In practice, with some exceptions, no actor has a panoramic view of the economy or knowledge of the distortions the public sector is supposed to correct (Kuznetsov and Sabel 2011). Instead the best that can be hoped for is a process of interaction or bootstrapping in which adjustments are made and some equilibration between demand and supply is achieved. This is consistent with an incrementalist understanding of development in which successful countries begin with simple and then move to more complex tasks, building the capacity to implement them as they go.

How should we think about such a process? The closest analogy is ‘social learning,’ in which performance is ratcheted up over time through trial and error, or as an evolutionary system in which there is variation or mutation to extend the range of possibilities, selection on some criterion of efficiency (‘what works’) and replication of selected variants (‘scaling up’). This requires a control or error-detection mechanism, defined as a set of institutions imposing discipline on public goods provision (Amsden 2001). It also requires a ‘sensor’ to detect and define the macro-context of the process to be disciplined, an ‘assessor’ to benchmark performance against objectives, an ‘effector’ to trigger behavioral changes on both public and private agents involved in the process and a communications network to ensure a transparent information flow. The key question is what institutional form these might take.

Why might learning not occur?

Why might this learning not occur? First, it might be that the system of policy-making fails to generate alternatives to the status quo. This could be because insufficient connections to the outside world exist to allow the inward diffusion of new ideas, or because of excessive bureaucratic centralization. Second, it might be that viable alternatives exist and are known of but are not selected. This might be because they are not in the interests of those actors, private or public, with veto power over the selection process – particularly in the case of ‘market interventions’, such as subsidies or trade protection whose benefits, unlike pure public goods, can be appropriated by individual firms or groups of firms.

But it might also be that there is insufficient information about the costs or benefits of interventions or that the connection between program evaluation and the setting of budget priorities is not properly articulated.⁴ Robust evaluations are difficult, time-consuming and expensive. Finally, there is a scaling problem. Learning requires effort and people’s intrinsic motivation to engage in it tends to be unsustainable. Also, much of the knowledge required to implement policy effectively is tacit in the sense that it is much more easily transferred via direct personal contact among practitioners than through documents.⁵

⁴ For these reasons, even in OECD countries, ‘ex-post evaluations tend not to feed back... into policy-making – they feed back into better implementation and, more rarely, budget decisions.’

⁵ Clusters of firms often engage in labor pooling to preserve the physical proximity and specialized knowledge of workers even when low demand may not allow them to hire (Criscuolo, 2002). It is also why firms with high returns

	Possible explanations	Possible remedies
Failure to generate alternatives	Intellectual rigidity; bureaucratic hierarchy; interest-based capture	Decentralize problem-solving; specify ends not means
Failure to select	Opacity of outcomes; interest-based capture; misalignment of cycles	Ensure ex ante transparency of objectives; require risk-sharing
Failure to replicate	Inadequate capacity or financial resources	Draw on diasporas; build capacity & networks

Serial versus parallel experimentation:

Experimentation can take place either in series or in parallel. Serial experimentation is akin to a depth-first search strategy, in which a smaller number of promising leads are pursued, while parallel experimentation emphasizes breadth first, pursuing multiple leads at the same time (Ellerman 2004). A similar distinction exists in biology between reproductive strategies that emphasize the quantity of offspring, pursued by organisms that have little control over their progeny’s chances of survival, and those that invest in a smaller number of offspring and in which the parents have greater influence over the environment in which those offspring are reared. In research terms, the former would be represented by a single large research laboratory; the latter by a community of scientific researchers, working in small semi-independent groups but transmitting findings rapidly through the literature and ratcheting up a common base of knowledge.

The pros and cons of these strategies depend on various factors (Nelson 1961). On the one hand, running multiple simultaneous research and development efforts might appear wasteful of resources; on the other, there may be significant uncertainty over the nature and cost of the best way to reach an objective, particularly at early stages of an investigation or research effort.⁶ Which is more efficient depends on the degree to which the solution is known in advance. When a high degree of confidence exists, it may be appropriate to focus on ‘delivery’ through the hierarchical imposition of discipline. At the same time, such an approach risks what in statistical terms would be categorized as a Type I error – rejecting a true null hypothesis by narrowing the scope of search prematurely. When the degree of certainty is less, it may be

to learning such as high-end professional services invest significantly in electronic networks to connect people (Jordan, 2012).

⁶ As Jane Jacobs put it, ‘Development work is a messy, time-, and energy-consuming business of trial, error and failure. The only certainties in it are trial and error.... Indeed, development work is inherently so chancy that by the law of averages, chances of success are greatly improved if there is much duplication of effort.’ Jacobs, 1969.

more appropriate to take a broader approach. But this then risks the incurring an error of a different sort – a Type II error – by failing to reject a series of false null hypotheses.

These considerations have some relevance for the structuring of development projects, which must decide whether to intervene ‘wholesale,’ i.e. with a larger number of beneficiaries but indirectly via some central agency, or ‘retail,’ by working more closely with a smaller number of beneficiaries. At the same time, their importance should not be overstated. Factors other than efficiency, such as equity and monitoring costs, must also be weighed in deciding how to allocate development assistance; also, the parallel between engineering and institutional solutions is at best inexact. Institutional solutions are likely to be replicable only in proportion to the degree of similarity in starting conditions across units.

Experimentalist governance:

Both serial and parallel experimentation require an organizing structure. Experience suggests that such a structure consists of three stages, linked in an iterative cycle. First, broad framework goals and metrics for gauging their achievement are provisionally established by some combination of ‘central’ and ‘sector’ units.⁷ These criteria should be derived from policy objectives and public interest requirements and not technical solutions, though the targets they embody must reflect the involvement of technical staff if they are to be achievable.⁸ Second, sector units are given broad discretion to pursue these goals in their own way⁹, allowing for decentralized problem solving. The aim is to allow space for innovative technical designs and operational practices to be introduced on the initiative of project managers – analogous to the open call for proposals that distinguishes private innovation competitions.¹⁰ Third, and as a condition of this autonomy, these units must report regularly on their performance and participate in a peer review of progress.¹¹ This should include an accounting of all public funds spent on interventions. The review may lead to adjustments in the specification of objectives and provide agents with opportunities to learn from each other.¹² Objectives can change in line with an evolving social consensus; it is also legitimate that targets be adjusted to take account of circumstances beyond the control of implementing agencies. At the same time, however, the peer review process can also lead to sanctions for persistent under-performance.

⁷ Sabel and Zeitlin, 2011, p. 3.

⁸ ‘The idea that a strategy can be brilliant and its execution flawed is simply wrong.... It misses the fact that everyone, from the top of an organization all the way down to the very bottom, makes choices under constraints and uncertainty, and so it becomes impossible to draw a line above which strategy happens and below which execution does.’ (Martin 2010, p. 68).

⁹ Sabel and Zeitlin, 2011, p. 3

¹⁰ For example, over 120,000 individuals around the world volunteered contributions to the design of Boeing’s 787 Dreamliner airplane (Piller, Vossen, and Ihl, 2012).

¹¹ Sabel and Zeitlin, 2011, p. 4.

¹² ‘Oscar Wilde once observed that architects differ from doctors in that they cannot bury their mistakes. Master builders have produced awesome works of art since the dawn of civilization, whereas it took millennia for the medical profession to transcend quackery and superstition (Rodwin, 1994, p. 213).

Building the constituent blocks:

The effectiveness such a system depends both on the quality of the constituent units (line ministries/agencies, service providers) and the articulation among them. It matters both how common decisions about how to allocate public resources are made (who exercises authority in consultation with whom and on what information) and how the executing agencies are staffed and organized.

The former is generally the responsibility of a ‘backbone’ organization, whose function is to broker between parallel problem-solving attempts between these agencies, monitoring their successes and failures and intermediating information about what works and doesn’t work and to communicate the results to political principals. The most commonly-cited examples include the Economic Development Board in Singapore or MITI in Japan, but a similar role can be played by planning ministries or implementing units within them.¹³ Typically these have been relatively small compared to the ‘sector’ units they coordinate, in some cases reflecting the need to concentrate scarce human capital at a point of maximum leverage but also because the sort of knowledge in which they specialize flows more easily from person to person in smaller organizations.

Backbone organizations: functions and examples

Function	Example
Sponsor parallel experiments and help monitor their outcomes	NDRC in China; EDB in Singapore, MITI in Japan, etc.
Serve as the source of creativity and dissemination of external ideas	IDR in Mendoza; Pemandu in Malaysia.
Maintain contact with politicians and strategy-setting	New Deal coordinating agencies in the US; Grand Councils in Europe post-1945.

Source: Jordan 2012.

Historical experience suggests that the composition of such backbone organizations or reform teams can differ significantly. They may be ‘elitist,’ following their own rules for recruitment, career advancement and remuneration (as in Taiwan) or ‘integrationist’ and subject to general civil service rules (as in Botswana). They be composed largely of in-country nationals or embed foreign technical experts, including in some cases secondees from the World Bank and other donors.¹⁴ However, common features emerge. The most effective backbone organizations have been multi-disciplinary, combining expertise in engineering with economics and other social sciences.¹⁵ They have also kept constant though selective

¹³ There is a direct parallel with private sector management findings that innovation initiative is best organized as a partnership between a dedicated change management team and a separate group that handles ongoing operations, the so-called performance engine (Govindarajan and Trimble, 2010).

¹⁴ In Mauritius, strategic engagement with foreign experts and technical assistance gained center stage in May 1970, when the World Bank seconded one director and few other expatriates to establish the Economic Planning Unit of the Mauritian Government and formulate the first Five-Year Plan. (Criscuolo 2007).

¹⁵ This ensured joint recognition of the need to support the development of efficient product and factor markets (advanced by ‘the economists’) with that of acquiring the project execution and manufacturing capabilities to compete internationally (advanced by ‘the engineers’) (Criscuolo and Palmade 2008).

links with external sources of technical expertise, sometimes through units whose sole function is to scan and produce digests of foreign research.

Managing the relations among them:

Then there is the management of the relationship between the backbone organization and the executing agencies (or 'performance engine'). As also emphasized by the literature on strategic management (Govindarajan and Trimble, 2010), 'conflicts between innovation initiatives and ongoing operations are normal and can easily escalate. Managers of the performance engine seek to be efficient, accountable, on time, on budget, and on specifications. An innovation initiative is exactly the opposite: it is, by nature, non-routine and uncertain.'¹⁶ These incompatibilities create an inevitable us-versus-them dynamic. Leaders must counter conflicts by constantly reinforcing a relationship of mutual respect. Antagonizing executing agencies can be disastrous. They are, quite simply, bigger and stronger and will always win in an all-out fight. For this reason, even the best innovation leaders need help from high places. They must be directly supported by political leadership that can override the short-term interests of line ministries where necessary.¹⁷

There are several ways of thinking about the relationship among these actors, embodying different combinations of discipline and autonomy. The traditional economic approach conceives of policy implementation as a string of principal agent relationships between taxpayers, politicians, line ministries, service providers and beneficiaries of those services. At one end of the chain are voters or other political constituencies; next come politicians, line ministries, specialized agencies, private service providers; and at the other end are the firms that consume these services to generate employment. Conflicts are regulated by contracts or agreements in which financial and other external incentives align the interests of principals and agents and deter or punish renegeing or shirking. This approach emphasizes the structure of the relationship among those public institutions and private actors involved in the chain of implementation.

A second approach emphasizes the professionalization of civil servants and others responsible for policy implementation. In this view, the principal obstacle to effective implementation lies not in the incompatibility of interests between principals and agents but poor information or uncertainty over the relationship between means and ends. Incentives matter but imposing them from outside can be counter-productive. Instead the emphasis should be creating a sense of calling or 'mission,' and allowing room for discretion on the part of work teams subject to their achieving overall performance goals (Tendler 1997).

A third perspective synthesizes these two approaches and leads us to the notion of 'diagnostic monitoring'. Unlike conventional monitoring which assumes infallibility on the part of principals, diagnostic monitoring allows that they can err in specifying means and ends and that careful attention to

¹⁶ Govindarajan and Trimble, 2010, pg. 81-82.

¹⁷ In the 1970s Botswana addressed this conflict by requiring planning officers to report both to their line ministries and to the Planning Officers Cadre of the Ministry of Finance and Development Planning (Criscuolo, 2007).

the problems agents face and the accommodations they devise can provide important clues about such error and confusion. But it does not exclude the possibility that agents may fail in their responsibilities. Persistent underperformance, as measured against the achievements of other agents in like position, results in penalties.

Successful models of iterative problem-solving exhibit at least two other characteristics (Jordan and Sabel 2015). First they depend on nested cycles of monitoring, in which successive layers of management reinforce and discipline each other. Typically, these will consist of weekly, monthly, quarterly and annual rounds of review involving increasingly senior levels of management. Second there must be some mechanism for deciding when to escalate decision-making from lower to higher levels of management. The default should be for problems to be resolved at the lowest level possible. But when this does not occur, either because of legitimate differences in opinion or for less justifiable reasons such as information-hoarding or obstinacy, there must be some means of pressuring parties to find a solution. Such a mechanism may take the form of a 'bumping-up rule' or credible threat of higher level intervention, with the risk that uncooperative behavior will be exposed in a professionally damaging way.

Structured agnosticism in World Bank projects:

How effective is the World Bank at supporting experimentation? To what extent might it play the role of a backbone organization to its borrowers? In theory, an institution like the World Bank that oversees thousands of projects and employs several hundred research economists should function as a backbone organization par excellence. It is true that the nested monitoring arrangements that distinguish recursive models of institutional decision-making, such as those in PEMANDU, go well beyond the typical bi-annual missions that characterize World Bank project supervision. But an external agent like the World Bank might still play a backstopping role, analogous to that of the highest level of political authority, particularly when disbursement is conditional on results and where these results are defined at a sufficiently 'downstream' level to allow for the emergence of independent technical solutions. As an external agent, it might also find it easier to ensure that local interests are fairly represented or that relevant information is considered in decision-making.

In practice, however, there are structural aspects of the institution that constrain its support for open experimentation. Probably the most salient is the reluctance of its shareholders to delegate authority to management for setting disbursement conditions. This has the effect of stifling the variety of paths (or alternative 'results chains') that borrowers can follow during a project without the need to revert to the Bank for approval: experimentation may still occur, but it will be strictly bounded. A second factor concerns the institution's inflexible review culture: insufficient allowance is made for the risk or complexity of projects. Instead the same critical machinery is applied to a USD 10 million as to a USD 1 billion investment, even when the purpose of the former is exploratory. A third constraint lies in its reliance on lending as its primary financial instrument. Borrowers are understandably unwilling to take on debt to finance open-ended innovation. Together these mean that Bank may be better suited to

supporting the scaling of tested interventions, with adaption to ensure local fit, rather than the early stage innovation more commonly associated with venture capital.

Attempts to encourage learning and adaptation through Bank lending instruments date back at least to the late 1990s and were founded in a recognition that the standard project cycle approach was inadequate for tackling complex or open-ended development challenges. As a 1997 Board Paper put it: 'Many of the urgent problems of development are precisely those which tend to tax the conventional approach the most: more and more we face situations in which knowing *ex ante* what works is not possible. This is particularly the case when decentralized and participatory implementation is the goal, and when the range of possible interventions becomes very large and differs over time.' The so-called strategic compact led to the introduction of two new lending approaches: Adaptable Program Loans (APLs) and Learning and Innovation Loans (LILs).

APLs were conceived as means of structuring a long, large or complex engagement as a series of shorter phases, to allow for adaptation and learning. The Bank's own review in 2012 noted that they could be effective in supporting the piloting and subsequent scaling-up of interventions, with opportunities for evaluation data to inform the design of follow-up phases; they also contributed to providing a platform for consolidating otherwise disparate policy actions. However, their effectiveness was undermined by the insistence on the use of pre-defined triggers or conditions for moving from one phase to the next. Bank teams also found that because follow-up phases required Board approval, they did not lead to any shortcuts in processing compared to a series of standalone operation. Finally, there was a tendency to overestimate the durability of borrowers' commitments to a single objective across changes of government. In practice the sustained attention required for effective serial experimentation often did not materialize.¹⁸

An approximation to support for parallel experimentation is to be found in the Bank's lending to municipal governments and in a few other sectors, notably education and health. The design of these projects is quite simple, typically consisting of two components: (i) institutional development through technical assistance and (ii) infrastructure and service provision through financing of physical investments. Sometimes they have introduced a competitive dynamic by making disbursements of funds conditional on some measure of performance, whether defined in terms of minimum conditions or fully scalable targets. The 'performance-based grants' model can take either of two forms: 'retail' in which the Bank works directly with a smaller number of borrowers, generally six or fewer; and, more commonly, 'wholesale,' in which it works through central government with a much larger number of municipalities. In most cases, performance-based grants have been used as means of persuading subordinate levels of government to implement centrally-defined improvements in financial management and other basic public management systems. They have generally not been used to encourage the exploration of alternative models of service delivery or other broader development objectives (Steffensen 2012). Instead

¹⁸ An attempt to remedy their shortcomings was made in 2017, with the Multiphase Programmatic Approach which was again construed as a means of piloting approaches to test assumptions and fit, with an emphasis on feedback and adjustment during the project cycle, but with the proviso that successive phases be approved by management rather than the institution's board of shareholders.

attention has focused on the management of vertical relations between the center and individual units rather than on exchanges across those units.

Good Government in the Tropics, Continued....

The Bank has a long history of working with state governments in Brazil, providing a combination of financial and technical support to strengthen public sector management of social programs. Its experience in Brazil was influential in shaping the design of the PforR instrument, which reflected several of the lessons learned there, particularly an emphasis on flexibility in the timing and scale of disbursements and the need for credible, though not necessarily independent, verification of results. One of its two longest-standing engagements was in Ceará, the northeastern state whose experimentation with public sector reform under Governor Tasso Jereissati in the early 1990s was the subject of Judith Tandler's *Good Government in the Tropics*. I took over responsibility for managing the Bank's largest and most complex loans to Ceará in 2010 and three years later was responsible for piloting the first PforR operation in Brazil. Over the two decades since Jereissati's first taking office, the state administration had strengthened its capacity to plan, finance and execute investments in physical infrastructure and basic services, especially health and education. Our intention was to deepen the model of results-based management and to push sector agencies to work together more effectively towards common goals. Drawing on ideas developed by Charles Sabel and others, and focusing on three critical challenges in water quality, early childhood development and skills development, the Ceará PforR was a conscious attempt to operationalize a system of diagnostic monitoring in support of complex problem-solving. It would be comforting but misleading to suppose that such efforts continued without hiatus: in 2015, an incoming government dismantled the previous administration's management model and miscommunication between the Bank, federal government and state together with a national fiscal crisis provoked a moratorium on the use of the PforR instrument in favor of more traditional lending.

The Bank's relatively new program for results (PforR) instrument, though conceived mainly other purposes, may be better able to accommodate iterative adaptation than its other more traditional approaches to lending. By disbursing against results rather than inputs, it allows more flexibility in how those results are achieved; and by using the borrower's own rather than the Bank's systems for managing fiduciary and environmental and social risks, it shifts attention from compliance to technical problem-solving. This makes it suitable for tackling more complex development challenges, such as those whose solution is not known in advance and which require a process of evaluation and adjustment. It may also be useful in improving the management of public investment projects through learning by doing (as opposed to formal technical assistance).

Concluding remarks:

In reviewing the experience of the World Bank over the past two or three decades, we have been struck by the similarity of concerns expressed by those involved at different times in trying to reform it. One interpretation is that this implies a failure to learn. But a second more optimistic view is that such repetition is inevitable: human organizations are entropic and require periodic injections of energy to

remain viable. And in this, as in so many other respects, we are reminded of Hirschman's observations on the non-linearity of economic development and on the need to call forth and enlist resources and abilities that are hidden, scattered, or badly utilized if it is to occur.