StanfordSOCIAL INNOVATIONReview

Features
Scaling Science
By John Gargani & Robert McLean

Stanford Social Innovation Review
Fall 2017

Stanford Social Innovation Review
www.ssir.org
Email: editor@ssir.org

→ Science delivers innovations that can spark social progress. But the common approaches to scaling up these innovations are based on commercial endeavors in which the goals are growth, expansion, and maximizing profit. The Global South, however, provides a different way forward where the goal is scaling for the public good.

Scaling Science

BY JOHN GARGANI & ROBERT MCLEAN

Illustration by JAKOB HINRICHS

In early 2014, the Ebola virus began its devastation of West Africa, moving through countries, communities, and families with grim efficiency. Over the next two years, 60 percent of those infected with the virus died—more than 11,000 people. A brutal killer, Ebola renders its victims delirious and unable to cope on their own.

One of the hardest-hit countries was Sierra Leone, which had just 136 doctors for more than 6 million inhabitants. Almost immediately, it fell to family and friends to act as caregivers. Ebola killed them, too. In the worst-hit areas, the virus eliminated entire families. Those who fell ill started running off to die alone rather than risk infecting loved ones. Eventually, social gatherings were banned, schools were closed, and households were separated. Society and the economy ground to a halt.

The crisis was unprecedented. Since Ebola was first detected in 1976, each of the subsequent 27 outbreaks was stopped in less than three months—until 2014. Why did this outbreak last for two years and kill more than all previous outbreaks combined? A complete answer has yet to emerge, but two factors were critical. First, we lacked know-how. There were no preexisting, evidence-based solutions to combat an outbreak of this magnitude. Second, the context was pernicious. A variety of circumstances, including unprepared health systems at the national level and social disintegration, compounded the problem and destabilized even the most holistic solutions.

In these types of circumstances, the way we usually scale solutions is ineffective. The traditional approach to delivering interventions at scale starts with the assumption that we have reliable solutions and favorable contexts. When this is the case, as it sometimes is, we are urged to scale what works by efficiently allocating resources to organizations with evidence-based solutions. But as the Ebola crisis shows, this is not always the case. Many of our most pressing problems are the ones we have been unable to solve, perhaps for years, for decades, or longer. Most are not crises on par with an Ebola outbreak, but fixtures of the status quo. Issues that in the development sphere are often called "wicked problems." So, how do we scale when we don't know what works?

TOWARD A NEW PARADIGM

In the absence of reliable solutions, or when new or changing contexts reduce the reliability of existing solutions, scaling depends on innovation. Innovation encompasses the entire path to scale, starting with ideas that hold promise and culminating in impacts that matter. It centers on innovators who are connected to systems of diverse actors. It depends on a dynamic body of evidence that develops before, during, and after scaling. Scaling innovations is justified by assessments of risk made by those put at risk, including those being served, and implies that trade-offs and values are weighed. It entails much more than resource allocation.

We need a broader way of thinking about scaling that takes our uncertainty into account and can be applied to the full range of contexts in which innovators, impact investors, funders, NGOs, social enterprises, and governments are currently acting. We are witnessing such an approach emerging across the Global South.

One of the organizations that are involved in combating the Ebola virus in West Africa is the International Development Research Centre (IDRC), a Canadian institution that supports innovations developed by natural and social science researchers in the Global South. (One of the authors works for IDRC, and the other is a consultant who works with it.) Along with many local and international partners, IDRC supported efforts to combat Ebola—from long-standing support to public health innovation in West Africa to rapid response mechanisms, including the trial and scale-up of a new vaccine.

The science behind clinical trials and large-scale vaccination is well understood. With some variation, it is the approach to scaling championed by organizations such as the Campbell Collaboration, What Works Clearinghouse, and 3ie. This approach has merit, yet it was not appropriate for the Ebola outbreak in West Africa.

Without turning their backs on clinical trials and other accepted approaches to creating and scaling solutions, IDRC and its partners worked to end the Ebola crisis in a different way. Their effort is one example of an emerging paradigm of scaling that we call "scaling science." This new paradigm is based on a review of IDRC's work that has aimed to advance a scientific or critical approach to scaling.

The term "scaling science" purposefully embraces two meanings. The first refers to the objective of *scaling scientific research results* to achieve impacts that matter. We define research broadly as the starting place of innovation. It is how solutions to stubborn problems are generated. From this perspective, researchers are innovators, and innovators are researchers.

The second meaning refers to the development of a systematic, principle-based *science of scaling* that we believe can increase the likelihood that innovations will benefit society. The aim is to contribute to building a culture of critical thinking on the topic. All approaches to scaling should be questioned, tested, refined, and used thoughtfully. We have learned time and time again from innovators in the Global South that it is the careful combination of imagination and critical thinking that leads to meaningful change.

TRADITIONAL SCALING PARADIGMS

Most of what we understand today about scaling up social change has been borrowed from 19th-century industrial expansion, 20th-century pharmaceutical regulation, and 21st-century technology startups. We refer to these as the industrial, pharmaceutical, and lean scaling paradigms. While there is much that we can learn from these paradigms, they are insufficient for contemporary social innovation. They reflect an old mind-set in which organizations rather than impacts are scaled up, scaling is an imperative, bigger is better, and the purpose of scaling is commercial success.

The industrial scaling paradigm is premised on the need to produce and distribute many standardized physical objects at the lowest cost. The key is "operational scale," and it is achieved by exploiting the efficiencies of large-scale manufacturing and distribution. Its purpose is to increase market share and, if possible, secure monopoly power. Replication, franchising, and train-the-trainer models, common in the nonprofit sector, are extensions of the industrial paradigm.

The pharmaceutical scaling paradigm is premised on the need to capture the sole rights to an approved innovation. The keys are "authority to scale," in which the government grants an innovator permission to scale up a drug based on phased clinical trials, and "exclu-

JOHN GARGANI is founder and president of the program design and evaluation consulting firm Gargani + Company, Inc., and president of the American Evaluation Association.

ROBERT MCLEAN is senior program specialist in the Policy and Evaluation Division of the International Development Research Centre, an organization established by the government of Canada to support research for development across the Global South.

sivity of scale," in which the innovator is empowered through patents and trade secrets to deny others the right to scale up the innovation. The subsequent challenges of operational scale—the manufacture and distribution of a pill, for example—are often trivial in comparison. Current approaches to evidence-based programming, favored by many governments and foundations, draw heavily on this paradigm.

The lean scaling paradigm is premised on the need to grow fast in a competitive market. The keys are "rapid learning," quickly iterating product designs to understand what markets value, and "resource scale," securing timely funds to exploit what has been learned and grow market share. The lean development process—build a minimum viable product, bring it to market, learn rapidly from customer behavior, modify the product or pivot, and repeat—drives many of today's leading tech startups. Unlike pharmaceutical companies, these innovators do not require authorization to scale, only the support of customers and investors, and they often find exclusivity difficult to enforce. As with pharmaceuticals, the problems of operational scale are usually negligible, especially if the innovators are selling intangible goods, such as software as a service. This is the paradigm that social entrepreneurs and impact investors are often encouraged to follow.

These three paradigms are strategies for achieving commercial success, not social impact. They may, however, provide some guidance for social innovators who want to scale up impacts in certain areas. A developer of low-cost irrigation systems adapted for African sunflower farmers, for example, may benefit from adopting elements of the industrial paradigm in order to expand production. Advocates for changing an environmental protection policy will likely benefit from the staged collection of evidence as one does with the pharmaceutical paradigm. And an education or e-health software innovator may benefit from basing its development process on the adaptive and nimble elements of the lean paradigm.

The old paradigms are not wrong when applied to social impact; they are incomplete. A more comprehensive approach would focus on an alternative or additional objective—the public good. With the scaling science paradigm, we set out to develop a framework that does just that. Our hope is that it will encourage innovators to consider scaling from a broader perspective, with tools that are inspired by the vast and eclectic problem-solving experience of the Global South.

FOUR GUIDING PRINCIPLES FOR SCALING SCIENCE

Scaling operations, revenue, market share, financing, and other aspects of an organization's work are familiar concepts. Scaling in these contexts is synonymous with growth, and more is better. They are legitimate organizational purposes. But our interest is in scaling social impact. It is not synonymous with growth, and more is not always better. From the perspective of scaling science:

> *Scaling impact is a coordinated effort to achieve a collection of impacts at optimal scale that is only undertaken if it is both morally justified and warranted by the dynamic evaluation of evidence.*

Embedded in this definition are four principles: moral justification, inclusive coordination, optimal scale, and dynamic evaluation. When these principles are not explicitly addressed, the public good may be overshadowed by other purposes—in particular, organizational growth. Scaling science is built on these four guiding principles, which are intended to help social innovators navigate the path from ideas to impacts.

1. Moral Justification | Scaling is not an imperative. In fact, sometimes it is better not to scale. The first principle, "moral justification," balances the pressure to grow against a responsibility to others. Researchers may feel pressure from government, investors, funders, and peers to increase the use of their innovation or to grow their organization. But in making that decision, innovators also have a responsibility to the people affected by their innovation. And part of that responsibility is met by the way in which scaling is justified.

Jeffrey Bradach, cofounder and managing partner of Bridgespan Group, has similarly pointed out the need to justify scaling. He suggests that program directors ask, "Is replication reasonable and responsible?" and challenges them to base their answer on evidence of effectiveness.[1] His is fundamentally a technical question. To answer it, program directors interpret the results of prior research and evaluation. But how much evidence of what type justifies scaling? Moreover, who decides?

We suggest asking an alternative question, "How certain must you be that your innovation will achieve positive impacts and avoid negative ones before you scale?" It is a moral question because scale is a value-laden objective. To answer it, innovators—and those affected by the innovation—need to establish criteria for scaling that are based on "acceptable impact risk."

The people who are affected by an innovation are the ones who bear impact risk. They will suffer if an innovation fails to produce its intended positive impacts or unintentionally produces negative ones. If social innovators scale before they are adequately certain about the impacts of their solution, they impose too much impact risk. If they scale too cautiously, they impose too little. Social innovators should search for an intermediate, acceptable level of risk. (See "Three Levels of Impact Risk" at right.) Among the factors that help determine what may be an acceptable level of risk are the urgency of the problem, cost of failure, diversity of perspectives, availability of competing solutions, and likelihood of negative impacts.

Often, scaling must be justified in the absence of laws or professional norms that might regulate it. Had there been no Ebola crisis in West Africa, phased clinical trials followed by large-scale vaccination would likely have been judged appropriate. In this case, there would have been norms and laws regulating how the vaccine was scaled up. As the crisis exploded, however, human lives were at stake and the urgency of the problem grew. Accordingly, a riskier strategy was accepted.

Those bearing the impact risk, including medical professionals, community groups, and policy makers in West Africa, were the driving force behind that decision. There were no fully tested and approved Ebola vaccines. So the decision was made to move forward with a vaccine that had demonstrated early trial efficacy in Guinea. In addition, local and international actors devised an innovative strategy of inoculation inspired by the approach used to eradicate smallpox in the 1970s. In this approach, a relatively small number of high-risk people (family, friends, and caregivers of known victims) were identified using network analysis and vaccinated. In the absence of an Ebola outbreak, this strategy would have imposed too much impact risk. In the midst of a deadly crisis, the risk was judged to be acceptable.

2. Inclusive Coordination | The second principle of the scaling science paradigm is "inclusive coordination," the idea that innovators must develop relationships with those affected by the innovation and those that make scale possible. Most of the time, it is beyond the capacity of a single innovator or organization to substantially improve a social or environmental problem, no matter how bold its scaling objectives. Scaling impact depends on the partnership, collaboration, inclusion, and competition of many actors. The practical challenge that innovators face is how to coordinate the actions of diverse actors with multiple agendas and perspectives in a way that advances the public good.

The actors who make scale possible include investors, funders, policy makers, government agencies, and customers. Their participation attracts considerable attention because they control financial resources and power. There are many models for engaging them, and they play a critical role.

Other actors also need to be engaged—in particular, those affected by the innovation being scaled. We have fewer models for engaging them successfully, and their exclusion typically does not jeopardize access to resources. Yet, they are best positioned to judge whether the impacts achieved constitute success. This is why decisions about scale—the what, how, when, where, and why—must also include the people who are affected. Our review of the experiences of those in the Global South indicates that doing so increases the chances that an innovation will scale successfully.

Three Levels of Impact Risk

The decision to scale impact depends on an assessment of whether scaling imposes acceptable impact risk on the people who may be affected by the program, policy, or product. If an organization decides to scale when certainty is low, it may impose too much impact risk. If it scales when certainty is high, it may impose too little. Seeking the middle ground—acceptable impact risk—governs the speed at which an organization scales impact.
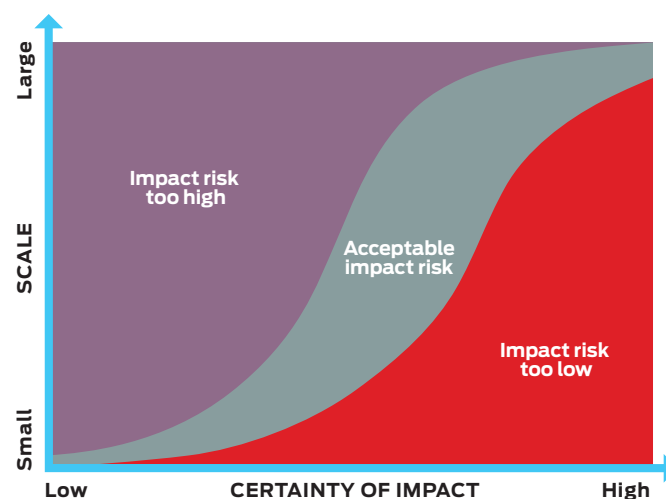
One approach to organizing actors is directed coordination, in which organizations and individuals come together, they agree on a plan of action, and one or more coordinate implementation. The Livestock Vaccine Innovation Fund is an example. It is a partnership between IDRC, the Bill & Melinda Gates Foundation, and Global Affairs Canada to develop, produce, and commercialize innovative vaccines against livestock diseases in sub-Saharan Africa and South and Southeast Asia. The fund coordinates the collaboration of diverse actors as they identify local needs, develop appropriate vaccines, and scale impact through uptake and implementation.

In the parlance of grantmaking organizations, directed coordination is often referred to as "collective impact."[2] Unlike many examples of collective impact, the development and scaling of livestock vaccines may require the coordinated entrance and exit of different collaborators at different levels of scale. Researchers doing the discovery science on vaccine candidates, for example, are rarely the same researchers who test vaccine efficacy in the field. Moreover, these researchers are even less likely to have the skills for commercializing and distributing animal vaccines to farmers. The actors playing a role in collective impact can change as scaling happens, and this requires a plan that incorporates anticipation, reaction, and facilitation.

An alternative approach to coordination is undirected. Here, coordination entails working together to develop organic systems—such as networks, markets, and professions—in which the independent efforts of many actors become self-organizing. The Community of Evaluators South Asia, a regional professional organization, offers a powerful example. It has helped develop evaluation systems in governments, universities, NGOs, and the private sector across eight South Asian countries. Their work has contributed to improvements in social enterprises in Bangladesh, sanitation programs in India, and the measurement of Gross National Happiness in Bhutan. The Community of Evaluators did not plan these outcomes in advance. Rather, they arose from the undirected interactions of the members of the organic system.

During the 2014 Ebola outbreak, local actors and international organizations coordinated their efforts. This allowed them to quickly scale up the vaccination program and help bring an end to the crisis. The combined work of scientists, health workers, and aid and humanitarian agencies is truly an impressive feat of collective impact. It is unlikely that organizations fighting Ebola would have achieved similar success had they made independent decisions.

After the Ebola outbreak subsided, many of these organizations and individuals continued to work together to address the underlying problems of public health that allowed the outbreak to reach the heights it did. They work together now to establish health systems in which preventive care and crisis response can be self-organizing in the future. In this instance, it was necessary to use both directed and undirected coordination to achieve sustainable impacts at scale.

3. Optimal Scale | The third principle is the idea that solutions to social and environmental problems have an "optimal scale," and rarely is it the maximum. There are trade-offs when scaling that typically make an intermediate level of scale the most desirable.

Understanding optimal scale starts with creating clarity about what exactly impact at scale is and how it will be measured. In our review of Southern innovations, we found that impact at scale is pursued well beyond minimalist constructs such as counts of beneficiaries. Other goals, such as improvements to a program's accessi-

bility for particularly underserved subpopulations or cost-efficiency gains, can greatly increase the overall impact of a program. At the same time, qualitative aims such as sustainability or satisfaction can also deeply improve people's lives. After all, it is entirely plausible that benefit for a population can be greater from doing very well on a small scale than doing less well on a large scale—and of course, vice versa. Small, slow, and beautiful or big, fast, and flawed—both can have their merits and their detriments. Embracing optimality requires being strategic about the level of impact we reach for and purposeful about its measurement. Aiming for "one million lives saved"—though bold imagery for a funder—can instigate unhelpful designs.

When assessing the level of scale that may be optimal, it is important to consider "scaling effects." When we move innovations from ideas to interventions in the real world, the scale of their impact is not a constant. As we increase our actions, the change in impact may be linear (additive) or nonlinear (multiplicative or exponential). It may also change qualitatively, becoming more desirable in type or nature. On the other hand, scaling may degrade positive impacts (diminishing returns), amplify negative impacts, and displace more effective alternatives. The way in which impacts change with scale—for better and worse, in linear and nonlinear ways, qualitatively and quantitatively—can mean the difference between success and failure.

Imagine, for example, if TOMS Shoes were to provide free shoes to everyone in a low-income country. The increase in the number of shoes provided would be linear, but the impact of the shoes would not be. While everyone who was poor and needed free shoes would get them, people who didn't need free shoes would also receive a pair, degrading the magnitude and quality of the impact. Moreover, providing free shoes to everyone would disrupt the local economy, hurting those who make, sell, and repair shoes. Not to mention the cultural impact of such meddling.

As scale increases, it may also change the mechanisms that produce impact. If we wanted everyone in a community to be protected against a disease, we would not need to vaccinate every person. This is because of a scaling effect called "herd immunity." As the proportion of vaccinated people increases, the probability that an unvaccinated person contracts the disease decreases in a nonlinear way. This is because there are fewer opportunities for healthy people to become infected.

In the case of the Ebola outbreak, herd immunity played a critical role. It was the reason why vaccinating only those at the center of the social network analysis slowed and eventually stopped the spread of the disease. Innovators on the ground understood this scaling effect and how it changed the mechanism of impact. They used that knowledge to establish an optimal scale for vaccination and a tailored strategy to deliver it. In other contexts, this level of scale may have been considered small. With Ebola, it saved resources, reduced negative side effects, and allowed actors to shift their focus to other areas of need.

4. Dynamic Evaluation | Impact evaluations assess the effectiveness of an innovation at a given level of scale. They assume stable cause-and-effect relationships, the kind commonly described by logic models and theories of change. In reality, impacts may become stronger or weaker, or qualitatively different, in response to a range of actions and scaling effects. To accommodate this, scaling science uses the principle of "dynamic evaluation," understanding how impacts change with scale.

To understand dynamic evaluation, consider a simple example of cause and effect. If you drive a car at a constant low speed (cause), it handles in a consistent and predictable fashion (effect). If you accelerate for an extended period of time, the car will begin to handle differently. This change is the result of scaling. Accelerating may help you reach your destination more quickly, or it may result in an accident.

Dynamic evaluation is, in effect, how we manage to drive vehicles at increasing speeds. We use a continuous and adaptive process of gathering, assessing, and acting on the signals we pick up from around us. It is dynamic because it can require changing approaches, frameworks, and theories as we proceed. In a car, if we hear a siren to our left, we turn our head and look out an entirely different window. In dynamic evaluation of scaling efforts, if a new innovation comes along that holds more promise, we may slow our scaling or change our designs for optimal scale. Dynamic evaluation can be viewed as a special case of developmental evaluation.[3]

The dynamic evaluation of the Ebola vaccine program started with a social network analysis. It was used to plan the targeted vaccinations. However, social networks were not stable because people moved to avoid hot spots and care for loved ones. This had the potential to reduce the effectiveness of targeted vaccines, so the analysis was updated frequently, vaccination strategies shifted with short notice, and rates of infection were monitored closely. To gauge whether the herd immunity had amplified the program's impact, evaluators looked for decreases in the infection rate that outpaced increases in the vaccination rate. These short-term results were used to decide how and where to roll out vaccinations. After the crisis subsided, the dynamic evaluation continued with more familiar summative evaluations and planning efforts to strengthen health systems.

TOWARD A SCALING THEORY OF CHANGE

To help innovators put these four guiding principles into action, scaling science aims to develop a new approach to creating a theory of change (a common component of evaluations and program designs), called a "scaling theory of change."

A traditional theory of change, which we call a "program theory of change," presents a plausible explanation of how a program is expected to achieve impact at a given level of scale. This level of impact is expressed as a static construct, often with graduated levels of similarly static activities, outputs, and outcomes to demonstrate a linear process of change that an innovation will travel to arrive at its eventual impact. A scaling theory of change, by contrast, presents a plausible explanation of how scaling is expected to change the way a program achieves impact as it scales. This is the key feature, and what makes a scaling theory of change different. In essence, it aims to capture the dynamism of innovation. It is intended to complement, not replace, a program theory of change. A scaling theory of change has three basic components: a path to scale, a response to scale, and partners for scale.

A "path to scale" is the sequence of stages through which an innovation is expected to pass as it scales. Any number of stages may be specified and named in a way that is most useful for the context. For example, a path may start with generating a promising idea that may produce a solution, followed by building the know-how to implement the idea, then applying the know-how to take action, and lastly expanding action to achieve impact at scale. This general path can be adapted to any type of innovation being scaled—for example, a policy, product, program, or practice.

Although the stages are sequential, an innovator's path through them rarely is. Advancing from one level to the next requires justification. Assessments of acceptable risk may result in a decision to move up or down one or more levels, or stay at the current level. Identifying the critical points where scaling should be justified helps ensure that scaling decisions are transparent, are based on relevant evidence, and include the people affected by the decisions.

A scaling theory of change also includes an explicit statement of how impacts are expected to change as the solution scales, called a "response to scale." These may include changes in the magnitude, quality, and type of impacts. In justifying these relationships, potential scaling effects should be taken into account. In most cases, an innovation produces a collection of positive and negative impacts, requiring judgments about optimality. Creating a visual representation can help stakeholders consider the trade-offs and identify an optimal point, or arrive at an acceptable compromise if they cannot agree.

The third component of a scaling theory of change identifies "partners for scale," and it describes the often-intricate arrangement and roles of partners involved in scaling up a solution. The principle of coordination plays an important role here. There are often two groups of partners, one collaborating on research and development, the other implementing and scaling the innovation. The work of partners within these groups requires coordination. Similarly, handoffs from the one group to the next, financial exits of private investors, and other transitions among partners may need coordination. Shifting arrangements of partners are unlikely to result in impacts at optimal scale unless facilitating organizations or self-organizing systems guide them.

LOOKING AHEAD

As the experience of the West African Ebola outbreak shows, scaling impact can be tremendously complex. There are typically many actors, changing conditions, and limits to our knowledge. This complexity can be found in the midst of a crisis and in ordinary, everyday life. With scaling science, we try to make the complexity of scaling more navigable. We do this by drawing from a vast diversity of experience of Southern innovators.

One way to think about scaling science is that its principles can help innovators create a map to guide their work. We can't chart precise directions for each and every scaling journey. As we travel, conditions change, and so too should our route, our speed, our means of transport, even our destination. But a map, built from the experience of others who have crossed the same terrain, can help us to plan a journey and evaluate a position.

We encourage innovators to consider what it means for impact to scale and how scaling decisions should be made. We invite inspired individuals and organizations to contribute to our understanding of scaling in general and scaling science in particular. How else can we hope to scale the science of scaling? ∎

NOTES

1 Jeffrey L. Bradach, "Going to Scale," *Stanford Social Innovation Review*, Spring 2003.

2 John Kania and Mark Kramer, "Collective Impact," *Stanford Social Innovation Review*, Winter 2011.

3 Michael Quinn Patton, *Developmental Evaluation: Applying Complexity Concepts to Enhance Innovation and Use*, New York: Guilford Press, 2010.