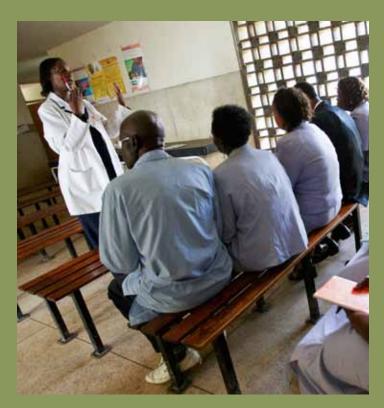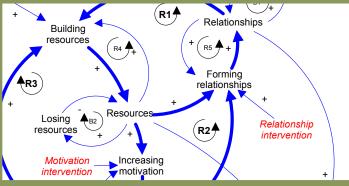# Achieving Lasting Impact at Scale

Part Two: Assessing System Readiness
for Delivery of Family Health Innovations at Scale

The thread we are following

There are several references in this synthesis to findings from an earlier convening held in Seattle in November 2011, *Achieving Lasting Impact at Scale, Part One: Behavior Change and the Spread of Family Health Innovations in Low-Income Countries*.

Findings from each of the convenings in this series are informing a dynamic publication, *How to Achieve Lasting Impact at Scale*, that provides a summary of the main messages as they are understood after each gathering.

These publications can be downloaded at www.dartington.org.uk/scalingimpact.

Achieving Lasting Impact at Scale
Part One: Behavior Change and the Spread
of Family Health Innovations in Low-Income Countries

A convening hosted by
The Bill and Melinda Gates Foundation
in Seattle, November 1-2, 2011
Synthesis and summary
by the Social Research Unit
at Dartington, UK

**How to Achieve
Lasting Impact at Scale**

What we can do
What we can't do
and
What we need to learn

Contents

Begin

How can we make a lasting impact at scale on maternal and child health? The answers to this question are at once tentative, ambitious, and urgent. They create a conversation of many threads. The thread I am following – and by putting these words on paper, inviting you to join – began in 2008.

That's when the Bill and Melinda Gates Foundation started to consider how it could contribute to Millennium Development Goals Four and Five, the goals to reduce child and maternal mortality and to achieve universal access to reproductive health.

By taking an interest, the Foundation placed itself in an unenviable position. Like other major funders and intermediary organizations, the Foundation discovered that, in the face of insufficient knowledge about how to create lasting impact, action with confidence was impossible. At the same time, in the face of eight million preventable child deaths each year, action with speed was essential.

Bill, Melinda, and their staff did what would be expected of new students. They read. They talked to experts. They lived in the places where child and maternal health was most at risk.

They began to understand that improvements to global family health were a product of two broad forces. There is science, which creates new knowledge that may help stamp out disease. Then there are people: those at risk at of ill-health, those with skills to prevent or treat illness, those working for local, national, and international organizations to get effective practices into routine use. These are real people with all the passions, constraints, accords, and disagreements of everyday life.

When I was invited to facilitate a series of exchanges to help those engaged in the challenge of achieving Millennium Goals Four and Five, I began to picture the opportunities produced by the confluence of available knowledge. Standing in the path of that confluence, ready to act, are people like Bill Gates who can digest systematic reviews and get to the nub of the problem. There are people like Melinda Gates who connect with mothers working within their limited and precarious options to keep their children safe and well. And there are people like the Gates Foundation's CEO Jeff Raikes looking for those "big bets," the investments most likely to have the biggest pay-off.

True, these are caricatures of well-rounded people. But what happens when every Bill Gates, every Melinda Gates, and all the Jeff Raikes in the world join a growing conversation about the Millennium Goals challenge?

I see my task as bringing together people who would not otherwise meet and encouraging them to talk about things they might otherwise set aside. In this case, the conversation includes scientists who analyze how systems change, and public and private sector experts working to scale up programs that help to save the lives of babies and mothers in poor countries.

I am looking for a new form of inquiry, one that is demanding and difficult, seeking cooperation from people who have separate and sometimes competing interests. But by brokering the ideas, stories and data of diverse experts – people who have dedicated their lives and careers to improving maternal and child health – new knowledge can emerge.

By the time I came to this thread of the conversation, its focus was clear. How can impact be scaled? Too many innovations, even the ones that are inexpensive, uncomplicated, and proven to prevent ill-health, have simply not reached their potential markets in Asia, Africa, and Meso-America. Upstream, we can spend our time coming up with new ideas. Downstream, we can try new and existing practices, processes, and policies, and see whether they sink or swim. This synthesis, like the other publications in the series, navigates the waters where the river meets the sea.

> By brokering the ideas, stories and data of diverse experts – people who have dedicated their lives and careers to improving maternal and child health – new knowledge can emerge

In the first *Achieving Lasting Impact at Scale* publication, I synthesized the findings from a major convening in Seattle in November 2011. The convening provided the space for a dynamic, exciting, sometimes haphazard discussion that framed what was to follow.

Complex as the discussion was, we found common ground in several areas. The focus in Seattle was not scale; it was impact at scale. Few people would have left Seattle, or put down the synthesis of the convening, with a linear view of the world. The arrows that schematize the world of impact-at-scale bend left and right, zoom up and down, and

double back many times before they reach their target. Users of innovations are not passive recipients of a distant good; they are integral to any potential good, fundamental to its success and failure. Seattle began a search for the lost "L" in MLE (Monitoring, Learning and Evaluation), a hunt that will lead to a new palette of MLE tools suitable for the non-sequential world of scaling impact.

I encourage you to read that paragraph again. It is innocuous? Perhaps a little technical? Abstract, and a long way from the hard realities of global health, such as the umbilical cord infections that kill half a million babies every year? For all its innocuous, technical abstraction, that common ground from Seattle tips upside-down many policy, philanthropy, practice, and science orthodoxies. These common themes demand a different tenor for the conversation as it continues.

And so the next formal convening took place in La Jolla, California, in March 2012. If Seattle was grand, with more than 200 participants and several major keynote addresses, La Jolla was compact. La Jolla relied more on a dialogue within a small, diverse, expert group. In Seattle we listened and commented in short bursts. In La Jolla, I sought a cool, thoughtful, mutual space for colleagues old and new to linger and get

into the detail of issues, the surface of which could only be scratched in Seattle.

The primary focus was *system readiness for impact at scale*. No one individual and no one organization can tackle alone the monsters of preventable infection, poor nutrition, treatable diseases, and better investments in health. Rather, differences are made by many thousands of individuals working in systems of organizations – and not just in one system, but many interlocking systems. There are systems around the catalyst, pushing out innovation; systems linking the catalyst to the delivery organizations; and systems around the governmental and non-governmental organizations that manage health care in most of Africa, Asia, and Meso-America.

I will say more later about what is meant by a system. For the moment, it has helped me to think of systems as living organisms that develop over time. They are made up of sub-systems, just as we are made of a heart, a liver, kidneys, brain and more. The functioning of the organism as a whole reflects the functioning of the sub-systems, and the functioning of each sub-system reflects the functioning of the other sub-systems, and of the whole.

The question we began to ask in La Jolla is this: when is a system ready to play its part in delivering impact at scale? For

the moment, let me say only that in the convening we began to re-form the issue. Reframed, the question is not whether a system is *ready or not*, but whether a system is *ready enough*. Readiness is a continuum, like a blood pressure reading that gives a sense of health without a 1-0 cut point. Systems are dynamic, moving in and out of states of readiness; they are made of many parts, some more ready than others. As a result, we need to be ready to test readiness again and again, looking for the places in the system that require extra attention.

Just as Seattle provided an opportunity to pressure-test new evidence and a model for thinking about scaling impact from Yale University, in La Jolla the conversation was greatly enhanced by the close inspection of evidence and models prepared by a team coordinated by Kaiser Permanente. It is testimony to the safe space for experiment that is provided by this continued conversation that Jim Dearing, Noshir Contractor, Peter Hovmand, and their colleagues were prepared to show a work in progress. Their ideas and tools undoubtedly aided La Jolla participants to understand more clearly the broader challenges of system readiness for impact at scale. ●

Who am "I"?

The benefits of bringing smart people together increase when their ideas are mediated in some way.

The role of the Social Research Unit at Dartington in this conversation is to broker, synthesize, and then broker again, trying to help all participants get the most out of it as it develops from Seattle, through La Jolla on to India and possibly beyond.

Our method for facilitating the convenings is described under "Brokering Knowledge" in the section on Talk. At its core, it comes down to gentle pushing and testing to find powerful shared ideas, to seek knowledge that is more than the sum of its parts.

We are curious outsiders, seldom promoting our own perspectives but always trying to channel others', trying to avoid cul-de-sacs and to keep the conversation flowing.

This approach is replicated in preparing a synthesis. For the most part we are reporting what has been said, but where there are

opportunities to develop the conversation further, we probe, just as we would in the convening, usually talking in the subjunctive ("if it were to be..."), imagining what could be true but never being sure, inviting someone else to develop the argument.

These doubting, hopeful observations are expressed in the first person singular to separate them from the collective voice of those taking part in the Seattle and La Jolla convenings.

So just who is the "I" you hear speaking in this synthesis? Of course, in reality, it takes several people to facilitate and synthesize a convening. For the La Jolla convening and synthesis, Dartington staff Michael Little, Dwan Kaoukji, and Beth Truesdale were helped by Tom Backer from the Human Interaction Research Institute, so "I" am one of we four.

Are we ready for "system readiness"?

Was I ready to faciliate the convening in La Jolla? Assuredly no. I never feel ready for these events. I always have the sense that I could have read more, talked more to participants, done more, thought more.

But I was confident of doing the best I could, and I like the fear that comes with a certain level of unpreparedness.

Was I ready to write this synthesis? That has changed by the day, from the fired-up energy in the wake of the convening, to the frustration of trying to balance so many complex ideas, to the doubt and hope that comes with releasing this synthesis into the conversation.

And, of course, as the "I" that facilitates and writes is actually several people, there have been times when one of us has been ready but another not. (Since we work in different parts of the globe, often at least one of us is asleep while the others are working!)

I tell this anecdote about my own readiness for two reasons. First, I hope the experience of feeling ready, not-ready, and ready again will resonate with many others, especially those who work in teams and larger systems. Second, it shows how, even in a simple situation, "readiness" is not simple.

Five ideas from the La Jolla conversation help to make a start at refining the definition. The first is that readiness is not dichotomous. We don't tick "1" for ready and "0" for not. Readiness is continuous, with few scores at either end of the spectrum. The question is not whether a system is ready or not, but whether a system is ready enough.

The second is that readiness is not constant. It changes over time, and for activities that involve partnership, the assessment must extend to the team.

A system that is ready today may not be ready tomorrow; a system that is ready to launch a product may not be ready to sustain it.

The third is equilibrium. In many situations, systems find a balance. They may not be positioned the right way up, or be facing in the right direction, but they may be settled. And if they are, then an innovation is going to disturb the equilibrium – by its introduction, and its growth, and even by its potential failure or decline. Along the way are lots of factors that can disrupt the equilibrium: the motivations of partners, their capacity to cope with new ideas and burdens, levels of trust within the system, and much, much more. Readiness may occur when the system is prepared for these challenges.

The fourth is fit. A system that is ready requires a fit among organizations within the system, and among the systems required to scale impact. This is not a jigsaw fit. It is more like the link between two joints in the body, which is often agile but sometimes painful and sore.

The fifth idea comes with the answer to the question, "Ready for what?" This conversation is about when systems are ready to scale impact. It sometimes gets reduced to readiness to deliver, or readiness to ensure that the innovation is used, both important intermediate steps. But we remind ourselves that the ultimate goal is impact, and impact is fundamental to defining readiness.

Talk

A convening is a small experiment in conversation. In La Jolla, experts with diverse experiences and views came together to talk toward a common language of system readiness. On a larger scale, *talk* represents the interactions among partners that are a crucial part of the complex social systems we seek to understand.

I have littered the synthesis with scale success stories (and at least one instructive failure). It is astonishing what is being achieved and the amount of partnership involved. Project Fives Alive! gives a flavor. In January 2008, in the area of Ghana in which the project began its work, only 10 percent of newborns were receiving care in the first or second day of life. Within 28 months, the innovation had boosted the rate to 70 percent. The project now covers all 38 of the districts of northern Ghana. This meant engaging with 185 sub-districts and 23 hospitals.

An ordinary partnership might involve a handful of agencies or a tie-up among a few businesses. Fives Alive! is the product of two international non-governmental organizations (the Institute for Healthcare Improvement and the National Catholic Health Service), the Ghana Health Service, the Ghanaian government, more than 60 primary health care organizations, those responsible for community health and planning, and the 20-plus hospitals. The intervention depended on 18 dioceses of the Catholic church, with evaluation by a team from the University of North Carolina collaborating with local researchers. And it is funded by yet another INGO (the Bill & Melinda Gates Foundation). As impact continues to scale, the number of partners will grow.

It would not strike most people as being unreasonable to take 28 months to get these partners into a room to make a plan, but within that time Fives Alive! is already delivering impact at scale. That achievement is partly due to continual attention to the readiness of all the partners in the system.

System readiness: what's a system?

At La Jolla, in the process of trying to understand how to assess system readiness – perhaps also how to encourage a system toward readiness – it became clear that the apparently simple, common words "system" and "readiness" both needed more definition.

At the simplest level, we can think of a system as a network of organizations working together to achieve an objective, which in the case of this conversation is impact at scale. I use the words "working together" cautiously; we all know that ordinary human interactions are subject to ordinary human frailties of jealousy, greed, idleness and pride, and sometimes worse.

It matters that these systems are networks of organizations and not networks of inanimate parts, as in a modern, computerized production line. Much system readiness expertise focuses on squeezing small efficiencies from the mass assembly of many parts sourced from many places into a single product. But the networks of organizations that can help scale impact on global health are social, not mechanical. They converse with each other; they talk. Each part adapts to changes by other parts of the system. The

system as a whole produces a reaction from the constituent organizations. All is in flux. The system is dynamic.

Having accepted the words "adapt" and "dynamic," we edge inevitably towards the word "complex." I resisted the descriptor "complex adaptive systems" when it was used in Seattle. I am slowly coming to understand why these words have become common parlance among experts in impact at scale, but I am still of a mind that, in the context of organizations working toward Millennium Goals Four and Five, "complex" and "adaptive" are two adjectives too many. All social systems are complex. All social systems are, by default, adaptive.

It matters that these systems are networks of organizations, not inanimate parts. They are social, not mechanical. They converse with each other; they talk; they adapt

They are also, to some extent, predictable. Although it is impossible to know for certain what a system *will* do, we can still make strong guesses about what it is *likely* to do.

Finally, in most of the scale impact stories I have encountered in this conversation, there are several systems, not one. These systems can be catalogued and named in many ways. Sociologist and physician Trish Greenhalgh, the "ghost in the room" in La Jolla, talks of resource systems, user systems, knowledge purveyors, and a range of other influences in the "outer context."

A different way of naming these systems draws on a distinction between what I would call *catalytic systems* that sit around the innovation and try to inject it into wider use

(the inventor, the primary funding source, the intermediaries supporting the initiative) and *delivery systems* that are responsible for diffusing the innovation (governments, NGOs, health systems, and other public systems). The components of each system adapt in response to changes in the others.

The task is to find the "sweet spot." This is the place where the complex conditions are as good as they can get; where there is enough equilibrium to provide a secure foundation, but enough dynamic imbalance to make change possible. Accurate tracking of this ever-moving spot gives the scale impact process the best chance of success in each stage of its development.

What is readiness?

I have already talked about how the definition of readiness began to evolve in La Jolla. The thoughtful definitions others have proposed set out a solid foundation for this task.

The Center for Global Development, drawing on the scale success stories it published in 2004 in the book *Millions Saved*, talks about the "capacity to affect change and implement strategies." The Kaiser team's READI report defined system readiness as "a set of partnering organizations that are assessed as being able to successfully deliver an intervention as determined by resources, relationships, motivation, and the contextual environment."

Wolfgang Munar paraphrased Trish Greenhalgh when he talked about readiness as "all players on board with protected time and funding." This formulation became a touchstone in conversation at La Jolla. But when they were asked, Emma Iriarte, Nana Twum-Danso, and Rajani Ved, all of whom have achieved considerable scale-up success, each said they had never reached this state.

Crucially, what became clear in the La Jolla convening is that there is no threshold of readiness, no point at which the partners can sit back and say definitively, "Yes, we are now ready to scale impact." Rather, a judgment call can be made at successive stages about the degree of readiness in a system as a whole, and about the parts of the system that are less ready than others.

Project Fives Alive!

About one in every 12 Ghanaian children dies before her or his fifth birthday. Project Fives Alive! aims to reduce this rate by two-thirds in a five-year period.

Progress is achieved not through a single innovation, but by moving from place to place improving health care from the first trimester of pregnancy to the child's fifth birthday. The motto is "start small, learn quickly and scale up rapidly," using methods developed by the Institute for Healthcare Improvement.

Local quality improvement teams convene every four months or so to agree common problems in health delivery and work out promising solutions. They then go out and test their ideas on a small scale, with the most successful being packaged into simple sets of procedures that can be easily adopted elsewhere in Ghana.

Launched in Tamale in July 2008, it is being implemented in four consecutive waves over a five-year period, moving from the North to the South of the country. Success is measured in improvements in the rate of perinatal care and neonatal care, and in the program's spread across Ghana's 170 districts. Care of sick children under age five is also a major focus of the project.

Photo: Five-day-old twins in a maternity ward in Accra, Ghana. (BMGF / Olivier Asselin)

Diversity

The discussion about the meanings of "systems" and "readiness" shows just how diverse are the partners needed for success, especially when the object is impact at scale on global health.

Contributions are needed from many people with many forms of expertise. In this context, diversity is not exclusively a matter of human rights or equal representation. Rather, diversity also becomes a deeply practical matter: a way to give ourselves the best chance of knowing what we need to know to improve maternal and child health, worldwide.

The convening in La Jolla is part of a process that tries to faciliate a conversation between two distinct groups of experts: on the one hand, the people I am calling the "scale experts," who have the rich and concrete first-hand experience of working to improve global health in low-income countries; and on the other, the scientists, who study scaling for lasting impact by dealing in abstract patterns, generalizations, and sophisticated analytical methods. And it aims to surround these people with other experts from the private and public sectors.

La Jolla differed from Seattle in that it gave more air time to the scale experts, those like Emma Iriarte, Nana Twum-Danso, and Rajani Ved, who are working on major initiatives like Fives Alive!, Salud Mesoamérica, or Saving Newborn Lives. Most scale experts live and work close to the people they seek to help, but not all do. Extraordinary people are achieving extraordinary things by combining an office

"It works, but it isn't used"

That phrase motivates many people in this conversation on scale.

If something works, why isn't it used? If a course on the subject were being taught, Oral Rehydration Salts (ORS) could be a case study.

For decades, there has been an effective and apparently easy solution to killer diarrhea: pre-packaged ORS sachets containing salts and sugars ready to be mixed with water can reduce diarrhea deaths by 93 percent. By the 1980s and 1990s, ORS was being manufactured commercially and UNICEF was committed to supporting its distribution via international NGOs.

But sustained spread of this simple, efficacious, cost-effective treatment has been miserable. After an initial scaling of impact in the 1980s and 1990s, during which time deaths from diarrhea halved, market penetration of ORS has fallen away. UNICEF suggests several reasons for the decline, including cuts in funding for social mobilization campaigns and practitioners' unwillingness to believe that such a simple therapy is "state of the art." And because the older formulation of ORS stopped dehydration but did not reduce diarrhea, many parents believed that it did not work.

Photo: Saru Auji lost her first son to diarrhea. She did not know about ORS at the time. (BMGF / Toni Greaves, Nepal, 2009)

life in London or New York or Washington, DC, with extensive time with their partners in "the field."

Success in scaling impact on global health relies fundamentally on the scale experts. Their kind of knowledge comes from engaging daily with real politics, personalities, compromises, strategies, and logistics. Through the personal experience of failure and success, their knowledge emerges as a form of intuition, a sense of what is possible and plausible, something like the knack of a skilled mechanic who knows by touch when all the parts of the machine are in order.

A second source of expertise resides in the academy. In the context of La Jolla, the bedrock was a systematic review by Trish Greenhalgh on the diffusion of innovations in service organizations. I have often heard that Greenhalgh's review was a touchstone during the Bill & Melinda Gates Foundation's early efforts to move their investments downstream towards and beyond delivery. Her work, synthesizing the wisdom of several research traditions on the spread of innovations in the Global North, took the Foundation a long way down the river. But the nearer they moved towards the Global South, to Bihar and Ethiopia, the less they felt able to rely on Greenhalgh's tools for navigation.

So the Greenhalgh work is being supplemented by the ongoing analysis of the Kaiser team, who have reviewed literature directly relevant to the task of improving global health and interviewed more than 100 scale experts around the world.

One product of the Kaiser team's work is a mental map of factors associated with system readiness. Thanks to their research, the list of factors was long when we arrived in La Jolla; and during the convening, participants added many more based on their own experiences. The taxonomy of factors (page 11) is my attempt to gesture toward what we have learned to date. These clusters of factors aim to show how complex, numerous, and multi-faceted are the foundations of system readiness, according to scientists and scale experts. For the funder, policy maker, or practitioner, this taxonomy can only be a starting point for the real questions: which factors are most important, when are they important, and how does each factor affect the others?

Indeed, the questions about the relationships among system readiness factors prompted the Kaiser team to engage specific scientific expertise: Noshir Contractor on "agent-based modeling" and Peter Hovmand on "system dynamics." ABM and SD, with their acronyms and complex mathematical underpinnings, can seem like "spooky science" to the uninitiated.

Of course, not all rigorous science relies on numbers – and some of the most interesting scientific moments are those when quantitative and qualitative approaches engage with each other. Neil Spicer's studies using interviews generated ideas about power differentials between organizations, for example, that helped reduce the "spookiness" of the computer-based models. In return, by the end of the convening, the mathematical models were being used to connect ideas in new ways, which is already informing Neil's next round of fieldwork.

The convening is part of a process that tries to facilitate a conversation between two distict groups: the "scale experts" and the scientists who study how systems change

So diversity in La Jolla brought together unusual combinations of stories and concerns. Diversity placed the trust-building challenges that are part of Rajani Ved's work in serving half a million Bihar villages, alongside a hypothesis, modeled by Noshir Contractor, that too much trust within a system curtails the spread of innovation. And diversity linked Emma Iriarte's efforts to connect eight governments in Meso-America and reduce inequalities in maternal and child health, with the Kaiser team's findings about how the motivation of system participants relates to the system's readiness to scale impact.

Brokering knowledge

A diversity of views is one thing. A genuine conversation among those with diverse views is quite another.

If making a conference more than the sum of its PowerPoint slides is a peculiarly modern dilemma, learning from presentations is no new problem. Mark Twain is credited with

saying that lectures can amount to nothing more than "a place where a professor's lecture notes go straight to the student's lecture notes, without passing through the brains of either."

Even when a balanced dialogue occurs, people often use the same words to mean different things. Progress is undermined by the chimera of agreement.

What is needed is a way to broker knowledge through conversation. This has been Dartington's role at the Seattle and La Jolla convenings, applying a method that I call Common Language. This method can be described in the simple terms of the good manners of listening, reflecting, being curious, and speaking clearly.

Common Language is based on the conviction that knowledge is not a product, but a practice. It is no use creating beautiful ideas if they are only tied up in neat bundles to set on a mantlepiece. Rather, convening, and creating what people might call learning communities, is actually one way of getting systems ready to scale impact.

Common Language also requires setting up a conversation that has the greatest chance of encouraging listening and learning. It involves inviting a productive diversity of talent and views to the conversation. It calls for an agreed focus, such as "impact at scale" or "system readiness."

Then, when the participants are gathered, a premium is placed on gentle testing of the meanings behind words and on allowing a shared language to emerge. The facilitation uses shared meaning to encourage constructive challenge, to link contributions with each other, and to test ideas rigorously but respectfully.

The arts required to make this method work are, in a sense, very simple: to listen, to define, and to connect. The conversation in La Jolla was diffuse and conversational. Everyone was invited to speak equally. Overlapping discussions were allowed to proceed without obvious solutions in sight.

There are two products of this form of brokering. The first is a series of words and concepts that cut to the core of the purpose of the conversation – words and concepts that can be commonly defined by everyone taking part. These words underpin the structure of this synthesis. They are

also articulated in the separate publication *How to Achieve Lasting Impact at Scale,* which grows with the learning from each convening. One example from La Jolla is the concept of "playing jazz" or "improvising": the idea that our systems become more ready for scale when we can allow more flexibility in our thinking about implementation, while still being aware of the overarching structure of the networks around us.

Even when a balanced dialogue occurs, people often use the same words to mean different things. Progress is undermined by the chimera of agreement

Common words form the basis for the second product, which is new actionable knowledge. There are two tests of Common Language: first, whether the sum of the conversation adds up to more than the individual contributions, and second, whether people do anything different as a result of the conversation. The strongest applications generally result from previously unforeseen connections between the strongest concepts underpinning the dialogue.

A vital point of brokerage in La Jolla was between the rich, deep, contextual knowledge provided by the scale experts and the emerging, exciting but as yet untried knowledge provided by the scientists. More back and forth, more considered conversation – in short, more Talk – is required to exploit the huge potential of this link.

Recurring themes, new themes

In La Jolla, I could see the strength of the themes discussed in the Seattle convening. The best ideas traveled effortlessly down the West Coast of America from Seattle, Washington, to La Jolla, California. For example, "context is king," as participants said again and again. It is not possible to listen to Rajani Ved talking about training 800,000 community health workers and contemplate everybody doing the same thing in the same way in every village served.

"Pull and push" also featured strongly in both convenings. Pull – the demand for

diffusion of an innovation by delivery agents and users, even when the catalyst is hesitating – is an important indicator of system readiness.

In Seattle we talked about how "collaboration is not necessarily altruistic," and the La Jolla days were littered with references to how the *sturm und drang* of doing something new contributes to the vitality of the process and has the potential to boost impact.

Alongside the familiar themes, new topics of discussion opened up in La Jolla. Perhaps even more exciting, there were new points of connection between ideas that first were voiced in Seattle. One new connection was the definition of complexity in the context of impact at scale. Another was the challenge (posed in Seattle by Don Berwick, Administrator of the Centers for Medicare and Medicaid Services) to create a new palette of evaluation tools. The new connections between complexity and evaluation will be discussed more in the coming pages.

These new openings will not deflect, however, from the primary objectives. In the coming pages, a shared understanding about what is meant by system readiness for scaling impact should become clear. Light should be shed on the challenges of getting systems ready for impact at scale, which, as will be seen, involves finding an optimal fit between catalytic and delivery systems. The lessons learned from the pressure-testing of evidence and ways of thinking prepared by the wider Kaiser team should also come through strongly. ●

READI: System Readiness to Deliver Health Innovations at Scale

Jim Dearing knows a bit about diffusion. He trained with Everett Rogers, and the two worked together for 20 years. Now he is in the business of helping get health innovations, such as cancer cures, to the people that need them.

When Jim was invited by the Bill & Melinda Gates Foundation to help them learn about diffusion in Asia, Africa, and Meso-America, he drew around him a strong group of experts, including Kendall Krause and Sarah Madrid, who have both worked in low-income countries.

Jim's team saw themselves as surveyors looking at a piece of land from several different vantage points in order to get the full picture. The objective was to identify the most important factors related to system readiness for diffusion and successful inter-organizational partnership that lead to impact at scale.

They sifted through over 12,000 publications before systematically reviewing over 200. They examined 30 decision aids intended to improve system readiness. And they interviewed over 100 experts around the world.

The first iteration of their work, which was continuing as the La Jolla convening got underway, was what they call the READI model, or System Readiness to Deliver Health Innovations at Scale.

From the many factors they identify as being related to readiness, they give priority to four sets. (The core of these factors are illustrated in the taxonomy on the facing page, along with more factors identified during the La Jolla conversation.)

(1) Resources: tangible resources (such as staff, money, equipment or technology, and infrastructure) and intangible resources and capacities (communications, leadership, and experience) that are needed to plan for and deliver a health intervention at scale.

(2) Relationships between organizations: the ties between partners trying to implement a health intervention at scale, typically government agencies, foundations, hospitals and intermediary organizations.

(3) Motivation: the psychological component of readiness, such as willingness and commitment from all the main actors in a scale initiative, from community health workers to funder executives and elected officials.

(4) Contextual environment: political, economic, social, technological, and legal forces that help or hinder systems of organizations seeking to take a health intervention to scale.
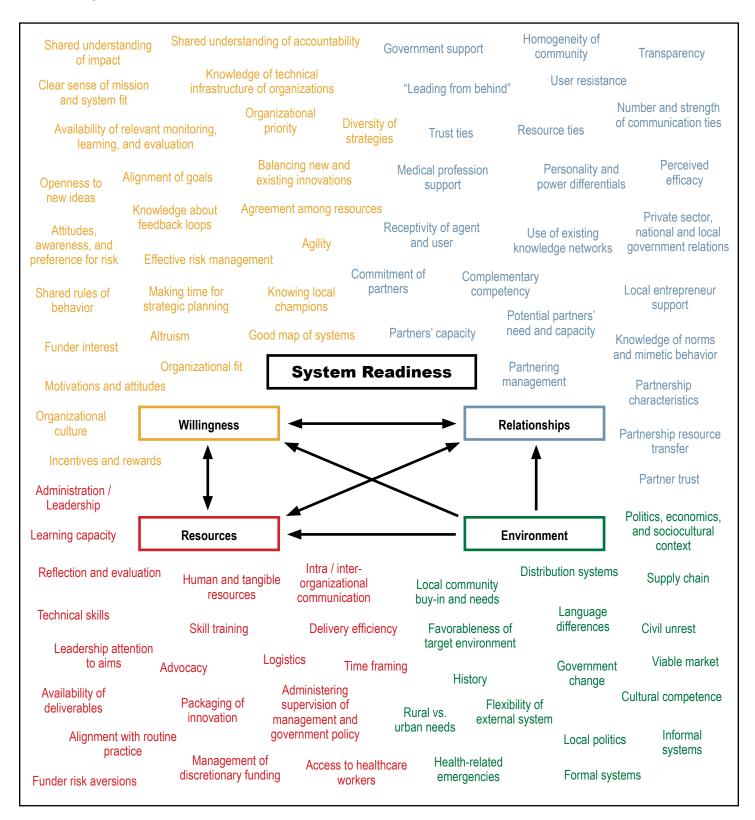
Presciently, from the beginning of the studies Jim Dearing engaged Noshir Contractor from Northwestern University and Peter Hovmand from Washington University in St. Louis to use their modeling expertise to make sense of the large numbers of factors making small but significant impacts on system readiness.

Factors that influence system readiness: a taxonomy

This taxonomy captures many, but not all, of the variables that were mentioned in the READI report and in the La Jolla discussions. It is intended to give a sense of the breadth of influences on system readiness, and to demonstrate that no initiative could hope to cover every angle. Of course, there are more factors than those recorded here. And it may be that in any single context, just a small number of factors acting in concert will be sufficient to make a difference.

Often, individual factors matter hardly a jot, but their interaction, multiplication, or ability to cancel each other out – which will differ from one context to another – matters a lot, and is the focus of much of the inquiry that follows. The diagram should not, therefore, be viewed as a source of data or a tool. It is a reminder of the volume of potentially relevant information.

Shared understanding of impact

Shared understanding of accountability

Government support

Homogeneity of community

Transparency

Clear sense of mission and system fit

Knowledge of technical infrastructure of organizations

"Leading from behind"

User resistance

Availability of relevant monitoring, learning, and evaluation

Organizational priority

Diversity of strategies

Trust ties

Resource ties

Number and strength of communication ties

Openness to new ideas

Alignment of goals

Balancing new and existing innovations

Medical profession support

Personality and power differentials

Perceived efficacy

Attitudes, awareness, and preference for risk

Knowledge about feedback loops

Agreement among resources

Agility

Receptivity of agent and user

Use of existing knowledge networks

Private sector, national and local government relations

Shared rules of behavior

Effective risk management

Making time for strategic planning

Knowing local champions

Commitment of partners

Complementary competency

Local entrepreneur support

Funder interest

Altruism

Good map of systems

Partners' capacity

Potential partners' need and capacity

Knowledge of norms and mimetic behavior

Organizational fit

System Readiness

Partnering management

Partnership characteristics

Motivations and attitudes

Organizational culture

Willingness

Relationships

Partnership resource transfer

Incentives and rewards

Partner trust

Administration / Leadership

Learning capacity

Resources

Environment

Politics, economics, and sociocultural context

Reflection and evaluation

Human and tangible resources

Intra / inter-organizational communication

Distribution systems

Supply chain

Local community buy-in and needs

Technical skills

Language differences

Civil unrest

Skill training

Delivery efficiency

Favorableness of target environment

Leadership attention to aims

Logistics

Time framing

History

Government change

Viable market

Advocacy

Availability of deliverables

Packaging of innovation

Administering supervision of management and government policy

Rural vs. urban needs

Flexibility of external system

Cultural competence

Alignment with routine practice

Local politics

Informal systems

Funder risk aversions

Management of discretionary funding

Access to healthcare workers

Health-related emergencies

Formal systems

Trish Greenhalgh: the "ghost in the room"

I n Seattle, the "ghost in the room" was Everett Rogers. In La Jolla, it was medical sociologist and physician Trish Greenhalgh. Many of the contributions to the conversation could be traced back to her systematic review on the diffusion of innovations in health service organizations published in the *Milbank Quarterly* eight years ago.

Greenhalgh's gift to the scale-up effort summarized a vast body of research on how innovations in the delivery and organization of health services could be spread and sustained. She and her co-authors examined 13 research traditions that all provided evidence on the question – but which rarely talked to each other.

Most importantly for our conversation, the review recognized that systems of organizations are frequently as important as, if not more important than, individuals in the scaling of innovation.

For Greenhalgh, system readiness happens when there is a healthy tension for change, a good fit with the innovation, and where the implications of change for the decision-making processes within organizations are well articulated and understood. Having more supporters than opponents is auspicious. Dedicated time and resources to support the early stages of scaling is essential. Capacity to evaluate progress also helps.

At the heart of the Greenhalgh model (below) are user, resource, knowledge, and change systems. This diagram has become a touchstone for many attempts to spread health innovations.

I met with Trish in a coffee shop in the East End of London on a wet April day not long after leaving La Jolla. She was full of helpful criticisms and insights, some of which I will come back to in the closing pages. She pointed out that Everett Rogers' "S-curve," one of our points of reference in Seattle, was hardly the only pattern for the diffusion of innovation. Adoption often follows other roads, such as Moore's chasm, Van de Ven's non-linear model, or Garter's hype cycle (page 31).

I was struck most by her distinction between consensus and accommodation. Implementation, she noted, needs flexibility and adaptability between partners who might accommodate each other but not entirely agree.

Model describing the determinants of diffusion, dissemination, and implementation of innovations in health service delivery organizations. Greenhalgh, Trisha, et al. 2004. "Diffusion of Innovations in Service Organizations: Systematic Review and Recommendations." *The Milbank Quarterly*, 82(4): 581-629.

Do

All our talking must be grounded in the world of doing. In the real places where the scale experts work and live, *doing* the day-to-day task of creating impact on health at scale means paying attention to the nuances of histories, personalities, politics, and power – and learning to "speak system" like a native.

Context is king

She said it in Seattle and again in La Jolla. Nana Twum-Danso started the convening proper with the words "context is king." I am going to push against this idea in the closing section, but the phrase resonated just as strongly in La Jolla as it did in Seattle.

Fives Alive! again serves as a good example. The project started in the North of Ghana. The North is poorer than the South, with a different ethnic composition and a distinctive political leadership. The political, institutional, community, cultural, and organizational contexts are very different in the two parts of the country. So at each stage in the development of Fives Alive!, a project that has been in the field for less than five years, a re-assessment of local conditions has informed the adaptation of the innovation, the strategy, and the logistics.

Government is among several powerful contextual forces important to the success of Fives Alive! In Ghana, the private sector is so small that partnering with government at all levels is essential, especially when reaching into rural areas. In projects like this, the distinction between the formal and informal public sector becomes important. In low-income countries, the range of formal health systems is limited. Health staff rely on community volunteers or community health workers. Some are paid a stipend, others not. It depends on the context. These details matter a lot to scale experts.

Scaling impact does not take place on a blank piece of paper. There is a history that precedes the innovation, and an inheritance that will survive long after catalysts around the innovation have packed up their bags and moved on to the next challenge. Progress depends on finding a fit with this context, this inheritance.

The personality of scale

Usha Kiran put a heavy emphasis on the human resources essential for impact at scale. She referred, of course, to the motivations and attitudes of the people who are putting innovations into practice. These motivations may involve altruism and a commitment to saving lives, but also include the ordinary needs to have a job, to feed one's family, to get ahead. Individuals tied up in the drudgery and worry of daily bureaucracy or miserable wages will not be able to deliver the best possible work. People also have particular ways of thinking and seeing the world, as we will return to later.

Like people, organizations and systems have "personalities." They care about impact. They can be passionate and trusting. They always need more money, which can be a state of mind as well as a hard reality. They have histories that shape their future development.

Many of the stories told in Seattle and La Jolla are about smart, skilled, passionate people trying to make a difference. These are stories about the head, the hands and the heart. They are stories about engagement, about doing. Many of the do-ers – many government officials, physicians, activists, and community workers – are motivated at least in part by a passion for saving lives. Some systems and organizations are fertile ground for the growth of passion; others are not.

Trust and security will help and, in the high-risk contexts in which impact on family health is sought, they take a long time to condense and a short time to evaporate. You may trust me because you trust my allies, long before I prove myself trustworthy – and this type of trust can be a mixed blessing. Al Bartlett and Nana made this point in talking about the World Health Organization. With WHO support comes credibility; but at the same time, credibility comes with a series of expectations that create limits.

Scaling impact does not take place on a blank page. There is a history that precedes the innovation, and an inheritance that will survive long after the catalysts have packed their bags

Like individuals, organizations and systems often don't really know when to stop, because so much focus is on the readiness to start (not even *if* to start but *when* to start). When an evaluation is put in place, it is done so with the assumption that outcomes will be good. As it was put in the convening, "It takes a certain level of humility as well as political independence to say honestly that 'we should not try this kind of scaling up right now – the timing isn't good.'" Systems and governments often aren't willing – or politically able – to do that.

Systems, like people, have their weaknesses as well as their strengths. There is a lot of second guessing about which horse to back. Rajani Ved used a vivid phrase in talking about the "fence sitters," those who watch and wait when two approaches are pitted against each other, looking to see which is announced as the winner. None of us likes to be on the "losing side."

Rajani gave a vivid example with respect to the innovation she has championed, Home-Based Newborn Care, which uses community health workers to reach a wide population of families in India whose health is at risk. But the context in India is changing. It is now a G20 country, a BRICS country, and so when the well-respected WHO and UNICEF come along and propose a model that focuses on doctors, those with the power to decide prevaricate and ask, "Why are we building community supports when in a decade or so we will have the resources to support a fully developed professional health care system?" Then one option is pitted against the other until eventually a compromise is found. In this case, the WHO/UNICEF model is followed by hospitals, while community health workers adhere to Home-Based Newborn Care. This division of territory would not be such a disaster, one may think, except that systems can collectively mull over these options for many years while public health suffers.

Ordinary people and ordinary systems can also behave badly. The functionary who seeks a bribe, the organization that operates against its own self-interest or the interest of its partners, the historical antipathy between parts of a system, the government systems that require that their officials be treated with due deference: all of this is typical in any attempt to scale impact.

Some people and some systems are simply not very good. Outside of the formal convening, the scale experts commiserated with stories of some government ministers with whom they had tried to work – the ones who admitted that they only realized what they should have been doing once they had been sacked, and the ones too self-centered to know they had ever done anything wrong. In system context, these are not just rogue individuals, but people operating within systems that allow or encourage incompetence or corruption, whether through perverse incentives, lack

Getting the insiders and outsiders alongside the insider-outsider

E mma, Nana and Rajani are insider-outsiders. They live in places where innovations are being scaled and work alongside the "insiders." But they work for, and interpret for, the "outsiders."

Insiders are the people who operate within the systems that will deliver and sustain the innovation. Outsiders include many of the people in the catalytic organizations, often major donors or intermediaries, that are promoting the innovation.

For insider-outsiders, much of their time is spent going back and forth between the parties, acting as a translator. It is not just a matter of translating English to Hindi but also of helping to make sense of the Western, aspirational, demonstrative and data-oriented culture that drives the catalyst. It is not just a matter of interpreting Twi for the English speaker, but of helping the donor to understand the reluctant-to-criticize, process-orientated, careful-adaptation, story-valuing culture of the delivery system.

Sometimes the work of the insider-outsiders is a matter of translating; sometimes it is a matter of finding adaptation that works for both parties.

Offline, the scale experts exchanged notes about the "sweet spots" for scale success – the places where they know enough about the local context with its unique history, successes, challenges and opportunities to be able to work out new solutions in partnership with the local people while bringing in new

ideas, knowledge, and possibilities that might otherwise be discounted by insiders.

Of course, just as insiders can get too comfortable with the status quo or over-pessimistic, outsiders can be over-ambitious and too pie-in-the-sky.

The right combination of aspiration and realism can be delivered by an insider-outsider working in partnership with leaders of the catalytic and delivery systems.

Reflecting on this idea after the convening, Nana Twum-Danso saw a point of contact between insider-outsider types and the Extension Agents that Everett Rogers described for the agricultural sector in the US. She used this model to some effect to train and coach insider-outsiders in Ghana, people who were inside their local context but who had been armed with new knowledge, skills, and experience about quality improvement and how it could help the system to achieve better health outcomes.

Photo: Head Nurse Pauline Ochola speaks to a group of community health workers, who work as volunteers within their communities, during a meeting in a hospital in Nairobi, Kenya. (BMGF / Olivier Asselin)

Saving Newborn Lives

Pediatrician Al Bartlett might well have thought he was due a rest.

But when he retired from the US government after a career dedicated to global public health, Save the Children asked him to lead their flagship program Saving Newborn Lives.

Al describes Saving Newborn Lives as "leading from behind" by supporting governments, health systems, and NGOs in 14 African and Asian countries.

There is no formula. Part of Al's task is to figure out if one exists, when innovations in each country are so distinct. In Tanzania, for instance, it met with success by encouraging the President to back the introduction of community health workers trained to improve the delivery of babies. The program is a collection of stories of scale-up success and failure.

Given this variety, what holds Saving Newborn Lives together is a public health framework, a commitment to data, strong community engagement, and a focus on getting government backing, because, as Al puts it, "in many of these countries that's the only manifestation of scale."

Photo: A Nigerian mother wraps her premature baby to her body. Her hospital was introduced to "Kangaroo Mother Care" by Save the Children. (NOOR / Pep Bonet)

of monitoring, rigid bureaucracy, or sheer disorganization.

So I began to get a sense of the personality of systems, their developmental trajectories, good and bad. The scale experts talked about the systems in which they worked collectively having attitudes and bearing grudges from past encounters, and using their resources to express their emotional commitments to the environments in which they operate.

Power

It was Neil Spicer's qualitative research in India, Nigeria, and Ethiopia that switched one track of the conversation onto power. In any set of relationships there are power imbalances. Any attempt to scale impact is in itself a powerful act that sends ripples and sometimes dangerous waves through previously placid systems.

Much of the preceding conversation about ordinary human frailties speaks to the way systems can exert power. In addition, the La Jolla conversation touched on four other sources of power – power that can promote scale impact or inhibit it.

First, there is the power of professions, particularly the medical profession. In the eight Meso-American health systems that Emma Iriarte seeks to reform, physicians lead teams at all levels, from local to hospital. Without a doctor's agreement, nothing can happen. In India and Africa, in the face of good evidence, physicians continue to resist the idea of community health workers giving injectable antibiotics. Professional norms can create a productive common language among practitioners; they can also become a powerful barrier to change.

Second, there is the power dynamic between program developers and those seeking to adapt and spread the program. Program developers often seek to protect their products, concerned that they will be implemented unfaithfully. So when Rajani Ved tries to adapt the intensive training model of Home-Based Newborn Care to reach 600,000 villages, she calls it "growth," while her opponents call it "watering-down."

Third, there are power imbalances within communities. Chris Galavotti and Robert Hausmann gave vivid illustrations of how simple knowledge-brokering techniques

could reduce inequalities and so produce innovative solutions and strong local support for those solutions. As Robert pointed out, in the context of community-based change, one organization may be limited in the impact it can have, but a network of organizations can create opportunities for substantial scale and impact. Power-authority dynamics create an adaptive tension that catalyzes these kinds of networks. Robert is preparing variations of these methods for use in emerging knowledge networks in India.

Finally, one elephant in the room in La Jolla was the power of major donors. Money and influence are critical to making change happen, but the gravitational pull from large funders can make life uncomfortable for those in their orbit. Tom Backer captured the irony with a quote from Woody Allen's book *Without Feathers*: "And the lion and the lamb shall lie down together… but the lamb won't get much sleep."

System pull

Joe McCannon, comparatively recently trained, said his education was rooted in the illusion of an innovation seeded and growing in a system in predictable ways. All of the La Jolla participants had learned, sometimes the hard way, that the real world is not like that. The complex dynamics of scaling impact are exaggerated by the fact that the change-maker and the innovation itself also will change the way systems behave, producing new contexts, new rules, and certainly new expectations.

Rajani Ved observed that simply injecting an innovation into a system is likely to produce an antibody reaction. Avoiding a destructive reaction involves making the injection at the right time, but also finding the right vein, the point in the system where an innovation can best be injected. But more than that, the serum has to be accepted by the body. As Rajani said, sometimes it is possible to figure out which elements of an innovation can easily be adopted by the existing system; then adding new elements piece by piece can lead to large cumulative changes in a complex system that might successfully resist the changes if they came all at once.

Joe and Rajani's observations propelled me back to Seattle, where we learned that "pull beats push every time." There we were talking about the pull from individual users; now we are talking about how to

create *system pull.* Traditional methods for scale-up involve a lot of push. A funder, a government, an intermediary, or a local NGO promotes an innovation, touts its virtues, and provides resources to support its use by health care workers, people, and communities.

The scale experts recognize the need for some push but right from the start they ask, "How much pull is there? Is it genuine? Will it come from leadership, middle management, the front line? Why would an organization want to do this innovation? Why would a community want it?"

For those on the frontiers of scale, genuine pull comes when the system comes to what we want for our own children; when the system comes to see the benefits of an innovation, and does so voluntarily, even enthusiastically.

For them the gold standard for scale-up success is that the system itself sustains its commitment to the innovation, and comes to take for granted the benefits that perhaps it previously acknowledged but never fully believed. The gold standard is to *systematically* deliver impact at scale. I was struck by a helpful tautology: reaching this standard means helping the system to do what systems do best – they systematize.

Vague generalizations about how scale experts think about pull and devolving right from the start, true as they may be, miss the full depth of the task

So, as Chris Galavotti reminded us, pull must be more than individual demand. Pull is about people in groups – in organizations, in communities, in systems – sharing information and jointly deciding what solutions might work for problems they've identified. When scale-up works, the whole system is pulling.

Now I am reminded of the AIDED model discussed at length in Seattle, and the most difficult part, the last D, to Devolve, which is the moment when we let the fledgling fly. The least controllable part of scaling impact comes when we let go. In system terms, Devolving means transferring power, ownership, and accountability for results.

Do you speak system?

The real-life task of scaling impact does not stop for convenings. From time to time between sessions, Emma Iriarte's phone would ring. Sometimes I could overhear her, speaking in a beautiful, clear Spanish, dealing with the latest challenge associated with Salud Mesoamérica. I found myself wanting to ask her "Hablas sistemas?", and if I could speak Hindi I might have asked Rajani the same. But I knew the answers to both. Emma does speak systems. Rajani switches as effortlessly between her native tongue of systems into impact as she does between Hindi and English.

I don't ordinarily see the benefit of dissecting the attributes of individual experts. I can understand that good leaders make for good schools, but how do I find a good leader and what can those whose management is lacking learn from the good leaders?

But there is something in the language and skills of people like Al, Emma, Joe, Nana, Rajani, Tom – and many others in Seattle and La Jolla – that strikes me as being intrinsic to scaling impact. And vague generalizations about how scale experts think about *pull* and *devolving* right from the start, true as they may be, miss the full depth of the task.

Here is one small example of what it means to speak systems. Emma said, "I know that for the oxytocin project to work, systems will want to know that it will work at the local level. They will need to know how much it will cost, where they will buy it, how it will be regulated, what rules need to be changed to allow it to be used, whether it needs to be included in the list of permitted drugs. If you don't do this, then nobody outside the pilot will be able to use it. Funding, planning, specific outcome measures – all of this is needed by the system before they can contemplate scaling up."

The value of this on-the-ground expertise goes beyond the basic fact that Emma knows the innovation and the politics inside out. Crucially, she thinks and speaks in the logic of the system. She locks on to the challenge of getting systems – markets, legal and regulatory systems, health systems, and

USAID Deliver

Sometime solid implementation of an old idea is the best way to go.

Ed Wilson works for USAID Deliver, and brings well-tailored but otherwise orthodox logistics solutions to the problem of delivering innovation in difficult contexts. By changing partnership arrangements in Paraguay, for instance, Ed managed to achieve a 30 percent improvement in the delivery of essential medicines.

After the economic collapse in Zimbabwe, the government ran out of money to deliver contraceptives. USAID and DfID had the money, but systems had little capacity to get the product to the right place at the right time. Wilson outsourced the delivery systems from government and offloaded the responsibilities for recording from midwives, two simple changes that boosted dispatch rates to 95 percent.

Ed finds that, as with other spheres of system readiness, bringing people together to achieve a common purpose is as important as the details of the logistic solution they agree.

USAID Deliver has developed numerous strategies to improve the availability of pharmaceuticals and other health commodities such as medicines for malaria, avian influenza, HIV and AIDS; family planning interventions; and laboratory reagents.

The primary focus of intervention is to strengthen local supply chains. Methods include improving logistics management information systems, streamlining distribution systems, identifying financial resources for procurement and supply chain operation, enhancing forecasting, and providing centralized procurement services.

USAID Deliver currently reaches 20 countries.

Salud Mesoamérica 2015

Long before Cristóbal Colón set off for Japan, bumped into the Bahamas, and changed Europeans' geography of the world, in the land that connects what we today call North and South America, a number of societies flourished.

However, in modern-day Belize, Costa Rica, El Salvador, Guatemala, Honduras, the southeastern part of Mexico, Nicaragua, and Panama, poor health plagues the poorest 20 percent of the population, mainly women and children under five years of age. High infant mortality, low rates of vaccination, elevated levels of anemia, and malnutrition belie the economic development of the Americas.

A collaboration of the Instituto Carlos Slim de Salud, the Banco Interamericano de Desarollo, the Spanish Government, and the Bill & Melinda Gates Foundation, Salud Mesoamérica 2015 helps each government in the region use local data to establish achievable targets to improve maternal and child health. Local incentives, close attention to the cultural context into which innovations are introduced, and a commitment to scale impact are the main features of Salud Mesoamérica 2015.

The program also uses results-based financing as part of its portfolio of tools. Advocated strongly by the World Bank, results-based financing refers to any program that rewards either the patient or the health care provider for an outcome, such as getting a child immunized or reducing mortality rates in a hospital.

Innovations focus on improved family planning, state-of-the-art intervention for obstetric hemorrhage, effective care for the umbilical cord, the introduction of micro-nutrients in the diet of under-twos and better treatment of diarrhea in under-fives.

Carlos Slim, widely acknowledged as the world's richest man, and a force behind Salud Mesoamérica 2015, has said that "health is an essential element for people to realise their full potential."

Photo: "Haciendo una linea de base." (Iniciativa SM2015, El Salvador, 2011)

governments – ready to systematize the innovation.

The language is supplemented by strong social skills. These experts carefully watch each person, each community, each organization, each system. They assess the mental map of each. They look for shared experience because they know it will help develop trust. But they also look for a fit between what is said and what is done, recognizing that actions speak louder than words. Those who are working on the frontiers of scale are catalysts both inside and outside the delivery system, and they rely on people with similar capabilities.

They are, to use Emma's words again, ready not to be ready. So their assessments are continuous, daily, monthly, quarterly, looking for the right opportunities to advance their initiatives. They start from the assumption that systems are never ready for impact at scale, that there will never be perfect measures of readiness, and that every moment we spend figuring something out, another child dies. So, to quote Tom Backer, sometimes the rule has to become "ready, shoot, aim."

Of course, like us, scale impact experts are ordinary people with ordinary frailties, characteristics shared by the organizations for which they work. Al Bartlett reminded the convening that powerful funders like USAID can make competing entities align with externally imposed goals, and that it is strong program managers who exert the US government's will. I have spent a lot of time talking about negotiation, finding a shared language and experience, but I left La Jolla better appreciating how systems can also value command-and-control, and how power can be used to create impact at scale.

Conclusions

What I most appreciated about La Jolla were the descriptions I heard of the world as it is, as opposed to the world as we would like it to be. The impact-at-scale world means politics, money, competition, and power. It means dealing with demotivated organizations and corrupt systems.

One option when faced with such a messy context is to step back and reflect: a natural inclination, but problematic. Wolfgang Munar expressed it best at the beginning of the convening. He spoke of the Bill & Melinda Gates Foundation, but his words will resonate with everyone involved in this project of improving global health.

Wolfgang said that we have insufficient knowledge to know how to scale impact, but with 22,000 children dying needlessly every day, there is an imperative to act. Systems have an inbuilt tension between stasis and change. The objectives of major funders grind against the goals of delivery systems, and we have to oil that joint. We use the best tools we can to measure our collective progress, but sometimes they get in the way.

Home-Based Newborn Care

While there has been significant improvement in the last two decades, more newborns die in India than anywhere else in the world – a combination of the large population and a high mortality rate.

Home-Based Newborn Care is one of many initiatives seeking to scale impact on newborn survival in India. The innovation targets preterm and low birth weight, neonatal infections and birth asphyxia.

The model adapts the community health worker approach with training of an Asha, or accredited social health activist. The Asha is involved from birth, through discharge from health facilities and the return home. Ashas prioritize premature and low birth weight babies. In addition to their direct preventative activity,

Ashas pass on health care skills to the mother of every newborn, and provide additional supports to those experiencing postpartum complications.

Home-Based Newborn Care has good evidence of impact (reduced neonatal mortality and the prevention of risky health conditions), and it has scale (the program is delivered in over 600,000 villages). Variations of the model being tested include incentive payments for accredited Ashas.

The innovation is partly funded by the National Rural Health Mission of the government of India and is administered at the state level.

Photo: A newborn and her mother at their Delhi home. (BMGF / Sanjit Das)

Helping Babies Breathe

A million babies die each year from birth asphyxia. As an innovation, Helping Babies Breathe could not be simpler: it gets people attending the birth to do the right things in the "golden minute" after birth.

It trains birth attendants in places where resources for newborn resuscitation – both knowledge and equipment – are scarce. The idea is to have at least one person who is skilled in neonatal resuscitation at the birth of every baby.

Since training materials prepared by the American Academy of Pediatrics were released two years ago, Helping Babies

Breathe has set up work in 34 countries, and has trained 33,000 birth attendants. It is backed by a number of partners, including USAID and Saving Newborn Lives / Save the Children.

In La Jolla, Wendy Taylor, who leads a new group at USAID focused on scale-up, attributed the rapid growth of Helping Babies Breathe to the fit with local health

systems, the ability of the training to demystify resuscitation, and partnerships with the private sector to design and sell affordable resuscitation devices.

Several La Jolla participants wanted further exploration into the factors behind this so far relatively young and unheralded success story.

Our ability to adapt to these tensions represents a major opportunity for better scaling impact.

I want to add two stories to make the point. Earlier in his career, Joe McCannon was responsible for a two-year health reform project in the United States. It was only when he got to the end of a six-month planning period that he realized the remaining 18 months were too short to implement those plans. He did not make that mistake again. Fast forward a few years, and Nana Twum-Danso, now Joe's colleague, is asking for advice on the Ghana project. She is a scientist by training, and although there is significant pull to move Fives Alive! forward,

she is not sure whether more analysis is needed before proceeding. Joe's advice is clear: give in to the pull. Reflect as you act.

These stories tell us that there is no blueprint for scaling impact. The task is a trade-off between the ideal and the possible. This means constant adaptation. It means innovating around the innovation, and being ready to improvise – something Joe called "playing jazz."

In this first part of the synthesis we have taken the advice of the Yale team: we went back to the user, became immersed in the contexts that we seek to influence, and looked at what people who scale impact "do."

As the previous pages testify, *talking* and *doing* are integral. But in the next section I will look from another angle. How do scale experts *think*, and what does this mean for what they do? ●

Think

How do the scale experts, the scientists, and other comers to the conversation think? Every way of looking at the world both obscures and reveals. Setting these diverse mental maps side by side may produce the creative tension we need to better chart the complexity of the scale landscape.

The objective, to improve maternal and child health, may be clear. But the mechanisms for achieving that objective are messy. How do Emma, Nana, and Rajani think about their task? How do they navigate the route that starts with a vague aspiration or at best a vision, passes through a major investment in an innovation, sees systems prepared and ready to deliver that innovation, and ends with many fewer mothers and children dying unnecessarily?

I have already talked about what this job involves on the ground, in the reality of the personalities involved, the language of "system," and what it means to *Do* system readiness. Now I turn to what we learned in La Jolla about what it means to *Think* system readiness – both for scale experts and scientists.

Scale experts use a map. It is a mental map that connects in some logical fashion not all of the landmarks and byways, but those that count as the most important points to pass on the journey.

Much of this work is taking place in poorly charted, choppy waters, but they are (in Emma's words again) ready not to be ready. There is no cartography team to work out the topography, hidden reefs, and dangerous tides. Indeed, in many instances, no technology exists to chart the geography accurately.

So it is necessary to make maps as the journey progresses. The landing points can be marked, and maybe the anticipated main roads and stopping points. This means identifying the primary organizations, examining how they relate to each other, in which system they reside and how well this system is functioning.

Joe McCannon talked about looking "broad and deep." These early assessments are mostly about the latter: exploring deeply, noting the details of the landscape.

Whatever the first sketches look like, the chart will have to change. Storms will blow away some of the terrain, shifting sands will create quicker trade routes. Anticipating change, constantly re-drawing the map, and being prepared to alter direction rapidly are all part of the scale experts' way of looking at the world.

Looking deeply is aided by the fact that many scale experts live in the places they are mapping. They will be there to find out if a highway turns out to be too unpredictable or expensive or inefficient.

It puts me in mind of Columbus setting off for China and Japan but ending up, to Spain's great fortune, in the Bahamas. Fifteenth-century sailors, coupling basic navigation tools with sensitive, experience-honed abilities to respond to the local condition of winds and waves, are perhaps not a bad analogy for today's scale experts.

Science has transformed travel, and human desire to pioneer has pushed at the limits of science. So what can science bring to help the frontierswomen of scale? And what can they bring to the science?

Where Emma, Nana, and Rajani go deep, science goes broad. The contrast may be particularly easy to see when comparing the scale experts to the scientists working with large data sets and complex statistical methods. But it is also the case that scientists working with interviews, case studies, ethnographies, and archives are often watching their small sample for what it can tell them about the broad, wide world.

There is no cartography team to work out the topography, hidden reefs, and dangerous tides. So it is necessary to make maps as the journey progresses, looking both broad and deep

So, in scientific fashion, the systematic reviews and interviews by the Kaiser Permanente team examined many attempts to scale impact in many contexts. The factors found to influence system readiness – resources, motivation, relationships, and environment – might apply to any large-scale change program.

The multi-dimensional models built by Noshir Contractor and Peter Hovmand seek to take the hundreds of factors identified by the Kaiser team, each of which makes a small but important difference to system readiness, and show how they connect, multiply, nullify each other, and act in concert.

These science-based, computer-based models thus produce a different mental map. It is an adventure that has been enabled by huge advances in technology, which make

it possible to calculate on a slender laptop a series of mathematical equations that would have kept Apollo 11 on the ground for several more decades had its mission depended on having such analytic power.

And despite their power, the models pressure-tested in La Jolla have uncertain limits. Part of the uncertainty comes from the fact that these models are charting a new course. (Remember that many expected Columbus to fall off the world he was navigating.) Partly it is that computers are very good at some things, but bad at others. The success of Amazon's "Mechanical Turk," in which businesses outsource small, repetitive tasks to workers around the world, is based on the fact that machines struggle with tasks that humans find trivial, such as matching similar photographs or transcribing song lyrics.

As Tom Henrich said, we are at the very start of building a GPS for scale – a satellite navigation system that can be carried around to improve the accuracy of the journey from innovation through system readiness to impact at scale. GPS was not perfect technology when it was first created, but it has become reliable. (Of course, if Columbus had a GPS, he might have made his way to China directly, but would not have gotten so fruitfully lost.)

So I am beginning to see this as two fields of vision. It may be a caricature – but I hope a useful one – to say that the scale experts are thinking deep, the scientists broad.

Deep to broad:
scale experts speak to science

Some of the factors implicated in system readiness for impact at scale are listed in the taxonomy (page 11). All of them, especially those that also form part of Kaiser Permanente's READI model, are valued by people working to scale impact. The sheer volume of data, however, makes it necessary to look for priority and connection.

I left La Jolla sensing that part of the skill of a scale expert is exactly this ability to spot priorities and connections, and to cut out extraneous noise – to be able to walk from one side of the room to the other without being distracted by the pattern on the carpet. They prioritize the major factors as they move from A to B.

Three simple, powerful questions

There are many frameworks, checklists, and other tools. There are complex mathematical models that help to connect up all of the moving parts. But back in the field, people charged with scaling impact often rely on simple but powerful heuristics.

Right at the beginning of the La Jolla convening, a note of recognition hummed around the room when Nana Twum-Danso referred to three questions she asks herself, repeatedly, as part of getting systems ready for innovation. She asks:

"Why are we doing this?" In low income countries, she elaborated, it isn't just that the health systems are ineffective at improving maternal and child health. There often is also a lack of alignment between health care and other systems. So she keeps asking, "why are we doing this" until she see some kind of congress of ideas.

"Who cares?" Of course, she observed, we need champions who will carry the work forward beyond the initial stimulus, and she looks not only to individuals who can be trusted but also to designated positions, to important job titles in key organizations, since the people she knows will come and go, whereas the employment structure of public systems is relatively static.

"Why should anyone else be passionate about the innovation I care about? Why would they add to what they are already doing?" The default position is for NGOs and government agencies always to be stretched, especially in localities where family health is most compromised. So they need a compelling argument for why they should add yet another service to all they are already doing – and they need to know not only if they will have to do more, but how much greater their burden will be.

Photo: Questions arise during an information session on family planning at the Makadara district hospital in Nairobi, Kenya. (BMGF / Olivier Asselin, 2009).

Each scale expert will have her own mental map, but I felt reasonably confident that they held common points of reference, such as a common starting point in the question, "What are we doing, and why?" The ultimate answer is "to scale impact on health." As in other parts of the world, in Africa, Asia, and Meso-America, systems are gradually moving (too slowly for most) from thinking about *outputs* to *outcomes*, and

from measuring *delivery and utilization* to charting *impact*.

Nonetheless, even when better health is the ultimate measure of success, the use of intermediate targets can help to drive progress. Wendy Taylor, for example, talked about the spread of resuscitation devices for newborns. Here, although the metric of most concern is the broad outcome of

The wood for the trees

All this talk of complexity got me thinking about not being able to see the woods for the trees, and that put me in mind of English painter David Hockney.

Hockney is experiencing renewed renown for his series of paintings of the Yorkshire countryside in which he was born and now lives. Watching him work is an education in perspective, on ways of seeing.

He plants a canvas in front of a wood and he paints. Standing over his shoulder, I can see on his canvas what I think I can see in front of it. But in fact I cannot. David will say, "You cannot see the top of those trees," and when I look I see that he is right. If I took a photo, the tops of the trees would be missing. But there they are on David's canvas.

He knows that my mind can project what is happening at the top, based on the patterns I see at the bottom of the wood. He puts on his canvas not what my eye sees, but what my mind sees.

I think this is like what is going on with people like Emma, Nana, Rajani and Usha. They cannot see the whole system (nobody can), but their minds can assemble a picture from their many examinations of the current reality, and from their long experience with other, similar systems. They are continuously taking snapshots in their minds, then joining them up to make a movie. They can see what we cannot. And just as a camera cannot compete with Hockney, so scientific machinery is struggling to compete with the gifts of these "scale-up artists."

Photo: David Hockney with Bigger Trees Near Warter, *2007, at Tate Britain. (©Guardian News & Media Ltd. 2009 / Graham Turner)*

better health, an important part-way goal is provided by the output measures that are more immediate indicators of delivery. Similarly, the push to encourage take-up of Chlorhexidine is, in the immediate sense, about utilization rates; only in a larger sense is it about health impacts.

Those responsible for scaling innovation will seek compromise when seeking an objective. A focus on the immediate destination (say, increased breast feeding in the first two years of life), will help all partners. But setting up different rest stops for individual members, such as more efficient delivery of health care, can also act as a strong collective incentive. And then they can focus together on the ultimate destination of higher survival rates and better health for young children.

Deciding where to go is the first question. Deciding how to get there is another. Scale experts seem to be continually asking themselves, "How will this get done?" With potential partners they might begin to sketch out the supply chain and start to compute the logistics, estimating the potential for wide delivery of the innovation. Edward Wilson had made significant impacts on family health in Africa and Meso-America by working with local partners to improve the delivery systems of contraceptives and drugs, often with simple alterations of relationships between public and private sector partners. For Ed, the innovation is to change system dynamics.

Partnership is implicit in the scale expert's perspective. This is a conversation that takes place in the first person plural. Rajani and her colleagues ask each other, "Can we do this?" The task by design is so huge that she wants to know who can help, and to what extent, at each point in the process. Most questions about capacity to deliver an innovation relate to several partners, including those within the catalytic system, pushing the innovation.

The first person plural also speaks to relationships. Organization A may have capacity to deliver the innovation, but giving them the responsibility may have negative effects on Organization B that then undermine the capabilities of Organization A.

Partnerships are initially sketched on a clean page of the planning book, but La Jolla reminded us that as the journey gets underway the paper gets crumpled, stained, and continually retraced.

Partnerships are about cooperation, but they also imply managing competition. From 30,000 feet, the world of scaling impact as represented in the models sketched by Noshir Contractor and Peter Hovmand is a series of quiet, orderly connections. On the ground, as Neil Spicer's qualitative research clearly demonstrated, there is the noise of the dog-eat-dog fight for survival among otherwise well-meaning NGO's.

In other respects, the deep maps of the experts and the broad maps of the scientists overlay nicely on each other. The Kaiser team put much weight on groups of factors that have to do with motivation and willingness. Local assessments of system readiness involve continuously asking questions like "why does this person want to help?" and "what is in it for this agency that clearly wants to be a part of this initiative I am putting together?"

The closer/deeper perspective bears more dividends when it comes to an assessment of "fit" between two elements. In Seattle, it was the fit between the innovation and users that figured most strongly, and La Jolla did nothing to diminish the importance of this connection. But the La Jolla convening

also brought to prominence the fit between the system that sits around the innovation (catalysts such as the inventor, the scientists who back it, the funders willing to support it, the international intermediaries that broker its introduction into a new setting), and the systems that operate in the context in which the innovation will be injected (government supporters, the public health infrastructures, formal and informal, the delivery agents and local communities).

I got the sense that proximity to the action and ability to readily re-reckon fit between their work and the wider ecology is a vital part of the scale experts' mental mapping. Their perspective has depth. It's as if they are looking with 3D glasses. They are making and editing a movie made up of a series of snapshots all taken close to the action.

This brings the conversation back to adaptation. The Latin root of adaptive is *aptus*, meaning "fit." For a scale impact expert, there is always an opportunity to find a better fit within a system, and between systems. In their view of the world, using their language, Emma and colleagues understand systems to be complex and adaptive and use these qualities to improve readiness for impact at scale.

Broad to deep: science speaks to scale experts

I can be pretty confident that the scientists take a different perspective and that they are looking broad before they look deep.

But even after lots of reading and two days of immersion, I am not sure if I know the models devised by Noshir Contractor and Peter Hovmand sufficiently well to be able to describe them succinctly. Maybe this doesn't matter. Thinking back to Seattle, I can hear Kristin Tolle, of Microsoft Research, saying, "You don't need to know how a smart phone works, just what it does for you."

A mathematical representation of a system is one type of model. It depends on the power of a computer program to run through thousands or millions of scenarios to work out how systems behave, and in this case to estimate how ready they are at any one time for the innovation.

Complexity is intrinsic to the world of systems. The mathematical models aim to make it explicit. Scale experts may

have a gift for being able to comprehend complexity that is way beyond the average person – beyond, perhaps, what they can even articulate. By contrast, to build even the most basic computer-based model, variables must be named and relationships must be defined.

And when the variables and relationships are well chosen and well clarified, then mathematical models come into their own. They do not forget a small but important feedback loop with which they are programmed, as a busy human might. And, of course, they have the advantage of speed; a computer can map out combinations much faster than a person can.

From a modeler's lens, it is also necessary to look at the whole system, examine how it responds and adapts, in order to highlight phenomena that would be missed by dissecting each element in turn.

Peter Hovmand and Noshir Contractor arrived in La Jolla with terms, relationships, and numbers based on the Kaiser team's research about how systems designed to deliver family innovations behave. These served as a jumping-off point, which the convening proceeded to dissect, challenge, and amend.

In terms of definition, for example, everyone in La Jolla agreed that motivation is a fundamental property of system readiness. The systematic review and interviews allowed the Kaiser Permanente team to define what is meant by "motivation": focusing on attitudes, building a sense of purpose for improvement, and, drawing on work by Joe McCannon, cultivating will and ideas. The modelers then had to further clarify how variables under the "motivation" heading can be operationalized.

After defining terms, modelers have to specify the relationships between each pair of elements. This can be seen visually in the graphical representation of the system dynamics model prepared by Peter Hovmand (page 26). This is a picture of complexity. It starts with the primary relationships between, for example, motivation and capacity. Does greater capacity boost motivation, or decrease it? This has to be specified. It then examines the feedback loops, differentiating between the positive loops indicated by "R," such as the way that an increase in motivation helps

The SUM of the parts

A three-step framework provides one productive way to think about scale.

In their work improving maternal and child health in hundreds of thousands of Indian villages, Poonam Muttreja, of the Population Foundation of India, and Rajani Ved use the Scaling Up Management (SUM) framework.
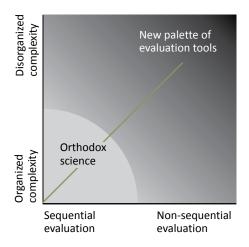
SUM is a three-step process. It begins with the questions about What (is being scaled up), Who (will do the work), Where and How. The second step focuses on the environment, advocating for the innovation, realigning and mobilizing resources. The third step deals with implementation and evaluation.

SUM tests every innovation to see if it is credible (based on sound evidence), observable (having impacts that will be seen by users), relevant (dealing with health issues that matter to local people), relatively advantageous (better than competing options), easy to implement, testable, and fundable.

A product of a collaboration between the Population Foundation for India and Management Systems International, SUM has been applied in half a dozen or more attempts to scale impact, including Home-Based Newborn Care.

Photo: A polio vaccination team in West Bengal. (BMGF / Prashant Panjiar, 2011)

The right tools for evaluating complexity

Our current science is ever improving its explanations of organized complexity. But for disorganized complexity, a new palette of tools is needed.

build resources, which in turn helps build more motivation; and the "balancing" or negative feedback loops indicated by "B," such as the natural attrition in motivation over time.

These models don't run without numbers. For the purpose of demonstration, the initial starting values were selected on the basis of limited data available to the modeling teams. And then the models began to produce food for thought.

In Noshir's agent-based model, for example, a high level of trust among organizations in a system increased the delivery of an innovation. But a *very* high level of trust was predicted to have a negative influence on delivery rates. Why? Perhaps because then cliques would form and new members would be excluded.

As another example, Peter's model made the important distinction between the delivery rate and the uptake rate for an innovation. It posited that delivery to target communities drives uptake but is moderated by fidelity, the degree to which the innovation is delivered as designed.

The modeler may have to define each term and relationship, but the modeling also allows for input from other experts – so much so that more of the La Jolla convening was taken up with Noshir and Peter re-working their models on the basis

of feedback from the rest of the group than was allocated to them describing their initial formulas.

After a single day of feedback, the scale experts had offered many important factors and relationships to add to the models, and had suggested which other, less critical factors could be set aside. And the La Jolla feedback also began to change what the models were calculating. At the start, both models predicted delivery – that is, they estimated the proportion of an innovation that would be delivered to the end user, or the amount of time it would take to deliver a certain percentage of the innovation. As the scale experts pushed back, the modelers began to look at how the models could calculate utilization rather than delivery, and impact rather than utilization.

Science and complexity

Listening to the way the scale experts and scientists made sense of the worlds they encountered put me in mind of the distinction between organized and disorganized complexity. Systems have both. The organized part is made up of the relationships that are largely predictable and orderly. We don't need research to tell us that innovation at scale depends on organizations with capacity working in partnership with similarly competent others to deliver that innovation; but that research exists. These data make sense to scale experts. Emma

Iriarte noted that government departments do not "think" rationally, but they do behave predictably. This part of the world is complex, but the complexity is organized.

Then there is disorganized complexity. This part of the world is made up of thousands of occurrences, any one of which is individually very difficult to predict. Power shifts in government are a good example. As journalists tell us, scandals involving ministers that force resignations and shift power may be predictable at an average level, and there will be conditions that make scandals and resignations more or less likely to occur. But which particular scandal, at what time, involving which ministers, and with what consequences – this is so difficult to anticipate that it can seem random.

Scale experts working in Africa, Asia, and Meso-America tend to plan based upon their knowledge of organized complexity, and adjust their strategy and logistics in the light of the effects of disorganized complexity. I suspect that, in their long-term planning, they often find it necessary to set disorganized complexity to the side; but that when the unexpected and the irrational does happen, they are willing and able to get down in the weeds and deal with it in order to keep the scale effort moving.

One of the values of complexity science is the ability to see patterns in what most people would regard as entropy. This is part of what the Kaiser team members are attempting to do as they model the relationships among the many factors that influence success at scale.

There is, I believe, a potentially important link here with a strong theme from the Seattle convening that returned in La Jolla. In Seattle, Don Berwick spoke of the "toxic" impact of traditional, sequential evaluation processes. To start with, the traditional tools are too slow for the speed at which decisions need to be made; while evaluators measure and weigh, the political debate moves on, and funding sources look elsewhere.

But even if current tools could be made faster, they would not be well suited to capture the non-linear, complex world of scale. Don Berwick called for a new palette of non-sequential evaluation tools suitable for the challenge of scaling impact.

It seems to me, going over all of the La Jolla discussions, that there is a relationship between organized and disorganized complexity on the one hand, and sequential and non-sequential evaluation on the other. It is best to illustrate this idea with a diagram, like the one on page 24. One axis captures complexity, ranging from organized to disorganized. The other axis deals with evaluation, going from sequential to non-sequential.

Increasing chunks of the world exist in the bottom left of the graph. A narrow set of well-developed methods – epidemiology, experimental evaluation, and systematic reviews – support relatively simple innovations that operate across social contexts. In this corner of the world, a ready array of metrics exists to estimate the unidirectional relationships between actions and impact.

But many of the Seattle and La Jolla participants operate in the top right-hand corner of the world, where context is king. Here, data may be diffuse, including stories and funds of shared experience. Potentially important variables may number in the hundreds or thousands. The relationships involve many feedback loops.

And in this world, impatience is a given. Whereas some types of action can be delayed while scientists finish sequential evaluation on organized complexity, when faced with disorganized complexity in the real world of keeping mothers and children alive and healthy, learning and doing become entwined. Joe McCannon spoke for all participants when he said, "I tell people who work with me, 'Go, get started, but learn actively and rigorously along the way.' I have action on a pedestal but it's not to sweep aside learning; it's a form of learning that's immersed in the activity."

So, in one sense, La Jolla turned into an attempt to build a GPS, a satellite navigation system, that combined the broad and deep perspectives from the range of expertise in the room. This GPS would predict for future travelers how long it will take to reach their destination, how much resource will be consumed, which is the most direct route and, maybe, what their options are if they run into traffic or roadworks. It would provide real-time data; it would take into account information that is currently overlooked.

How the private sector thinks about scale

Tom Henrich tipped his cowboy hat to bid me goodbye as he left La Jolla for another adventure. He reminded the convening that the private sector thinks about scale with the question: "Will customers buy what we offer?"

As a senior executive at Procter & Gamble, Tom's explorations have included new business creation in baby care under the Pampers brand. He noted that scale-up in the private sector starts with an idea, then moves to product development and prototypes, followed by consumer testing, technical testing, analysis to see who will buy, scale potential – so much preparation.

The focus, he said, is pull, which means, "Will consumers buy what we're offering?"

He gave as an example P&G's Children's Safe Drinking Water project, where the product is a handful of magic that cleans dirty water. A five-cent sachet will purify 10 liters. But when Tom got to Kenya, sales were terrible. The average shop was selling just three or four sachets a week. When he asked why, shopkeepers said the packets were too expensive. But in the same stores, two-cent lollipops flew out the door at the rate of 100 a week.

So Tom's analysis focused on what drives willingness to purchase. When the benefit isn't immediate, sales are harder to make (this is the great problem in prevention; people keep smoking, in part, because they die 30 years later, not right after the next cigarette). Other factors such as income, cultural and community norms, and ease of use play their part.
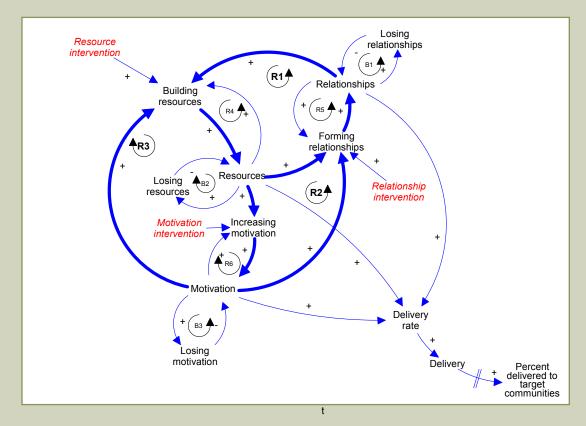
Tom's work for P&G will typically involve a system analysis to see what's happening with all these factors, looking for small changes that will reap significant rewards.

For private-sector explorers like Tom, willingness to purchase is estimated at every stage. Their experience has taught them that money has to change hands. People who tell a researcher that they will pay for a product don't always put their hand in their pocket when that product finally goes on sale. Tom says, "Always ask for money to see if people will buy, even if the price is low during the testing phases, and even if the product will be free when its distribution is scaled up."

Photo: Kamal Kant Jha's roadside paan and tobacco shop in New Delhi. (BMGF / Prashant Panjiar, 2010)

I began to see the convening as an opportunity to examine whether new scientific methods and new products from academe could help scale experts to make more sense of the disorganized complexity that surrounds them on a daily basis. I am anxious to follow Don Berwick's advice in Seattle and press on and build my non-sequential evaluation palette. The models discussed at length in La Jolla are strong candidates for the palette. The pressure-testing taught us not only about their value in helping to understand system readiness, but also what qualities will be required of future monitoring, learning and evaluation tools. ●

A system dynamics (SD) model of system readiness developed by Peter Hovmand and his team for La Jolla. During the convening, Peter changed the model as participants suggested new factors and new relationships among the factors.

What's in a model?

It is human nature to frame the aspects of our world, to set out its dimensions, to build a taxonomy, think about how the parts relate, and find a language that describes how it functions. Sometimes we can boil this down to a simple checklist that provides reminders of the things that need to get done, and in what order, to get the best result.

But many aspects of the world are complex and not yet properly understood. They resist simple categorization, and often we don't really know what factors to include or exclude. We can make checklists, but they are often incomplete or untried. Trish Greenhalgh's model of what makes for effective diffusion in health (page 12) illustrates just how complicated it is.

This world is dynamic. Many of the parts influence each other, and the situation at one moment is different from the next.

The way we think about these problems was changed by the creator of system dynamics, Jay Forrester. Early in his career at the Massachusetts Institute of Technology, Forrester helped a multi-national work out why it had such a high turnover of employees. The company had assumed it was due to forces out of its control, but Forrester worked out that it was a product of the human resources system.

Stocks and flows are at the heart of Forrester's way of thinking, and indeed to many who "speak system." For instance, when it comes to children in state care throughout the world, for every 100 children in care at any one time (the stock), 65 will enter the system each year and 65 will leave (the fl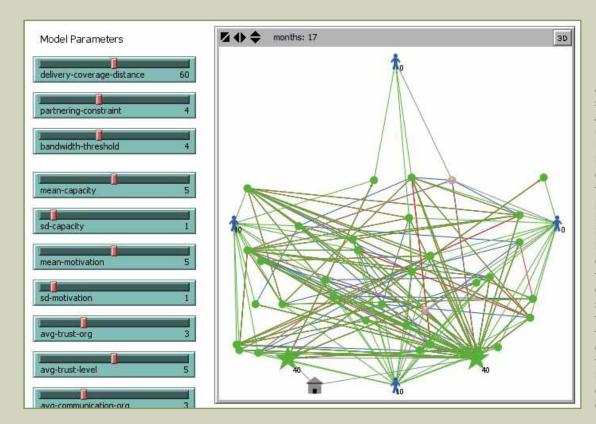ow). An imbalance in these figures usually indicates disequilibrium for the system and bad news for kids. People wanting to put things right usually look at the stock, but the solution is generally to be found in the flow.

Two other types of variable make a difference. Time is important. Taking longer to get a product to market, or keeping patients in hospital for a few more hours, will influence flows. And then there are the feedback loops (page 32). For example, a health administrator could demand that hospitals process more cases to clear a waitlist (an increase in the inflow), which demotivates nurses (decreasing the outflow). This increases the waitlist, which in turn results in even greater political pressure to admit more cases – and so the feedback cycle begins.

In the diagrams summarizing the systems dynamic (SD) model built by Peter Hovmand and his colleagues, we see terms familiar to his profession. A circle with an "R" indicates positive reinforcement, when one thing leads to another good thing; a "B" indicates negative reinforcement or "balancing."

In its purest sense, a model is an abstraction, a simplification. The ones in La Jolla were computer-based, but models don't have to involve either software or math. When Forrester sat in a factory in Kentucky in the early 1950s and did all his calculations by hand, he was sketching a model to represent how the real thing might behave. Harvard professor Gary King talks about a stick figure as a model of a human being. As he says, there is no such thing as a "true" model; only a "useful" model. The stick figure works very nicely for a game of hangman, or to indicate whether a restroom is for men or

Model Parameters

delivery-coverage-distance   60

partnering-constraint   4

bandwidth-threshold   4

mean-capacity   5

sd-capacity   1

mean-motivation   5

sd-motivation   1

avg-trust-org   3

avg-trust-level   5

avg-communication-org   3

months: 17   3D

A screenshot of a simulation run by the Agent-Based Model (ABM) developed by Noshir Contractor and his team shows the characteristics of the ties between each pair of organizations in the system – ties involving higher or lower levels of resources, communication, and trust. The bars on the left can be used to adjust the starting characteristics of the system – for example, to see how the eventual delivery level is predicted to differ if there is a higher average level of communication at the outset.

women. But it is not a very useful model of a human being if you are preparing to conduct heart surgery.

Forrester's challenge was complicated, but not that complicated, so a few doodles in a notebook linked to some simple math did the trick. That was in the mid-1950s. Then computers came along. Soon mathematicians were building computer models, stick men that danced with each other, including the marvelously named SIMPLE model: Simulating Industrial Management Problems with Lots of Equations!

Crucially, not only can a computer crunch bigger sums faster, it can simulate some of the random features of the real world. It runs its calculations thousands of times over and reveals the range of possible outcomes, as well as the probable outcomes. Statisticians call this Monte Carlo simulation. I may get lucky by betting my salary straight up on a single number of the roulette wheel, but if I do the same a thousand times over you can bet I will be poorer.

So when Peter Hovmand builds a system dynamics model, he is simulating the flows (the delivery of an innovation), the time it takes at each step in the process, and the feedback loops (if the warehouse is full of stock and the flow has to be diverted, this may set up a whole chain of failures).

A system dynamics model rests on the idea of equilibrium. If I think of a system as a balloon filled with water, and I press on one side, it will adapt, but once I remove my finger it will revert. Equilibrium is the natural state, but not necessarily the best state.

When Noshir Contractor builds an agent-based model, he is guided by the rules that govern relationships or "ties" between agents – in this case, adaptive organizations – in a system. The interaction and impact of these rules are too complex to see with the naked eye, but when a computer adds them up they turn into patterns that are at once beautiful and challenging.

In most other respects there is a lot of commonality between system dynamic and agent-based models. They both draw their data from many sources, including systematic reviews, qualitative interviews, and in La Jolla, from experts at scaling impact.

They both paint a picture of the world as it is now, and how it might be in the future depending on our actions.

They both demand that we establish clear definitions of variables and agree how, under different conditions, these variables will interact.

And they both tell us things we might not otherwise have known, suggesting small changes that leverage major benefits, for example that too much trust between organizations can be counterproductive or that increasing the motivation of organizations may bring more reward than improving their relationship to others in the supply chain.

In La Jolla, we saw these complicated pieces of machinery whirring away. It was exciting. But the thrilling part was sitting together, defining, refining and changing the premises of each model to reflect our understanding of the evidence or our particular expertise – and then watching how it changed the predicted success of our missions.

Dream

The dream is to build tools that will help the frontierswomen and frontiersmen of scale. Could the computer-based models that were pressure-tested in La Jolla – these "what-if tools" – be an early version of a GPS for impact at scale? Participants met the tools with hesitant optimism, and began to imagine their potential.

If I were to summarize the collective mood at the end of two intense days, after we were all bombarded with stories about scale-up success and failure, having examined the innards of models that are difficult to conceptualize, and having maintained a strong thread of conversation from Seattle, I think most participants were thinking along the following lines:

(1) This is very important. (2) There is a lot still missing from our analysis. (3) There are other good reasons to hesitate. (4) The lines of enquiry we are pursuing have a lot of potential, and we want this conversation to develop.

What is missing from the analysis

If it is hard to get a system to start, it is usually harder to get it to stop. When people talk about system readiness for innovations that can scale impact on global family health, a context that is dominated by the public sector, they generally mean "readiness to start." But most innovations fail, and there was a sense among La Jolla participants, unsupported by research but real nonetheless, that foundering initiatives are maintained too long. We learned that readiness was a continuous variable that required continuous review, but nonetheless the conversation stuck mostly to the pulling of the proverbial trigger and seldom strayed onto when to call off the race.

In the private sector, lack of money is the primary reason for pulling the plug on

an innovation. Tom Henrich, a veteran of multi-national private-sector-funded initiatives, created an antibody reaction with his observation that "viable market" was a factor to be considered in the system readiness mix. He observed that revenue plus the donor's outlay, minus expenses, must equal at least zero for an innovation to be sustained. Impact at scale requires a revenue source – whether from consumers, or from the public sector responding to demand from its citizens. From the standpoint of an organization like Procter & Gamble, market considerations are obvious data points for estimating system readiness, but they have received scant attention in this strand of the conversation.

Specific analysis is required on the question of competition from other ideas, priorities, and products. The private sector approaches a market with the desire to win and to see the opposition fall away, regardless of which product is "better" (think of the triumph of VHS over the technologically superior Betamax). In the public and philanthropic sectors, such rivalry can interfere with the public good. For instance, as Tom pointed out, an NGO that starts distributing a drug free of charge may undermine the low-cost solution that others have worked hard to systematize. Or the rival option may be a better bet for achieving the best outcome. At very least, much can be done to reach consensus across systems about how to minimize the negative and boost the positive effects of competition.

More work is required to consider the macro political environment into which innovations are released. For example, the increasing resources of emergent economic nations, India and South Africa especially, have to be factored into the system readiness equation. There will be greater state investment in family health. More centers of power – local, regional, or national – will be created to disburse money and manage investments. The new structures may be more efficient and responsive to local needs; or they may be more corrupt, with more fingers dipped into each additional pie. Much of the La Jolla conversation steered clear of the powerful influence exerted by central governments.

More work is required to consider the political environment into which innovations are released. In emergent economic nations, new structures may be either efficient or corrupt

There is also a need to better understand both the formal and informal components of public systems. The two models being pressure-tested rest in part on theories about organizational behavior. Historically, much academic work in this area has looked at legitimate organizations with reasonably well-functioning formal structures. Less is understood about the "under the radar" behaviors of organizations and the individuals within them who are seeking to get things done in the face of chronic red tape, inefficient use of resources, or corruption.

Finally, we talked too little about the unintended consequences of well-meaning actions. No matter how high the successful

delivery rate, an intervention that stokes local rivalries or creates lasting stigma will not help achieve the ultimate aim of improving health at scale. In the systems world, unintended consequences might be counted as "emergence": the way that complex results emerge from simple interactions, just as the infinite number of individual snowflake patterns emerge from simple rules of crystal formation.

Building a new palette of evaluation tools

The system dynamics and agent-based models discussed in La Jolla could be new colors on the new palette of evaluation tools that Don Berwick called for in Seattle – a new toolkit that is both faster and better adapted to charting impact at scale. What was learned from the analysis about what will be required of the new generation of methods?

As I said in *Think*, in the part of the world where sequential evaluation meets organized complexity, there are a few well-established methods with clearly agreed functions (such as epidemiology to establish potential causes of ill-health, or experimental evaluation to estimate impact). Where disorganized complexity demands non-sequential evaluation, a larger number of less well-defined tools are on offer. Few of these have been validated or have commonly perceived utility.

Wolfgang Munar referred to the models presented by Noshir and Peter as "what-if tools." They allow the user to draw on several sources of incomplete data – in this case, from systematic reviews, interviews with experts, and the body of knowledge assembled in La Jolla – and ask, for instance, "What if I were able to improve trust among the primary organizations in the delivery system?" The answer is not definitive. First, its predictions are only as good as the data and the assumptions that go into the model. Second, the answers are predictions about what will happen *on average*, based on a thousand or more simulations run by the computer; but we cannot run the world a thousand times, and there is no guarantee that our particular world will be "average." Still, by combining the best available data, the best guesses of the models may be better than the best guess of any single expert.

Indeed, participants were excited by many potential functions of the models. Could these tools be used to prime new projects to get them system-ready? Maybe they could give a push to stalled projects? There was broad agreement that the models might be used retrospectively to try to explain why initiatives like oral rehydration therapy had met with limited success; or prospectively to indicate what might be done to boost the potential impact of a new initiative.

So one unresolved question in La Jolla was like the old saying, "If all you have is a hammer, everything looks like a nail." What could these particular hammers be used to pound? Sometimes it works to develop a tool and then search for its function; other times, better results come from starting with the function (a task to be done, a question to be answered) and developing a tool to match.

Another unresolved question was, "Why two computer-based models?" Another way to frame the question is this: are these two methods looking for a question, or are there two questions that require different kinds of answer, or are these two competing ways to answer the same question? Jim Dearing suggested it was a matter of looking at the problem with two types of camera. The system dynamics model is like a wide-angle lens looking at the system as a whole. The agent-based model is like a telephoto lens examining the relationships among the organizations that make up the system. I could imagine artist David Hockney using both cameras and creating a collage; two types of camera might meet his need. So what would best help the frontierswomen of scale?

There was also much conversation about how to validate the La Jolla models, but it was not clear whether the limited data we have about scale-up successes and failures are enough to make these models robust. It may be that only by using a tool do we discover if the factors on which it rests (in this case, motivation, resources, relationships, and the environment) are distinct or whether in fact they are measuring the same thing. Whether the tool does the job will also depend on greater clarity about what output or outcome we are trying to predict.

There are many types of validation, each of which will demand invention from the non-sequential evaluation community. The tools that Don Berwick calls for may be radically different from current tools, but we will want them to match up to existing high standards of science.

Modeling the end of the world

How accurate are computer models at predicting the future? Peter Hovmand reflected on the work of the Club of Rome, a Swiss think-tank that modeled the sustainability of the planet in the early 1970s.

Their 1972 report, *The Limits to Growth*, caused quite a stir and sold over 12 million copies to readers no doubt curious about one of their predictions: that the world would run out of food and energy by the year 2000.

The systems being modeled by the Club of Rome are different from those that matter to scaling impact, but no less complex: the food system, the industrial system, the population system, the non-renewable resources system, and the pollution system.

There has been a rejoicing in the failure of this prediction almost as enthusiastic as the British pleasure in crowing over a mistaken weather forecast. But delving a bit deeper, several of the predictions in *The Limits to Growth* dealing with food, industry, and pollution have come true. Other predictions might also have come true, had world activities carried on with "business as usual" – or at least, as was usual when the models were made. Moreover, the feedback loops in the model set a standard for analysis.

Looking back over the Club of Rome report, I am struck by the modesty of the soothsayers. Like Peter and Noshir, they saw their models as providing indications of systems' tendencies given a set of assumptions, not as a crystal ball.

How far can we go with a checklist?

"**A**re you a left-ear surgeon or a right-ear surgeon?" Given the increase in specialization in Western health systems, this old joke is no longer so far from the truth, according to surgeon Atul Gawande.

On the one hand, specialization can be a helpful way to manage complexity. On the other, it can set the stage for mistakes, as in the case of the care of a patient where each expert working on the case knows something slightly different than the next, and where the cascades of health data are so massive that even computerized systems cannot always keep up.

Yet underpinning every job and every condition are simple procedures that, if done in the right order, reduce error. Gawande's book, *The Checklist Manifesto*, is littered with examples from medicine, construction, and the airline industry. It is a treatise on how to organize complexity.

Gawande illustrates the approach with a description of physicians' efforts to

prepare a checklist for the World Health Organization to improve surgery in diverse conditions around the world. At one level, failure of hospital care in South Africa bears no resemblance to failures in Asia, Europe, or the Americas. But a small number of routine changes, effectively implemented by health professionals who understand, believe, and can see personal benefit, can significantly reduce major risks to health.

The question for participants in La Jolla was the extent to which a checklist of some of the factors associated with system readiness could reduce risks to initiatives seeking to scale impact.

Photo: Polio vaccinators are trained in Kano, Nigeria. (BMGF / Prashant Panjiar, 2010)

of experience, gut instinct, and regularly updated mental maps. They were open to the innovations proposed in La Jolla, but they were keeping a sharp eye out for ideas that will help them do their job better rather than ideas that demand (without clear evidence of impact) that they do their job differently.

There is a need to delineate the boundaries around any tool. The models reviewed in La Jolla were seeking to address the question, "When is a system ready for innovation in global family health?" In practice, the models asked, "How long will it take under different conditions for systems to deliver innovations?"

A primary challenge for non-sequential methods will be for users to understand enough to be confident about what they are seeing. There is no point for a would-be scale expert to first become a computer engineer. At the same time, a major challenge for models like the ones produced by Noshir's and Peter's teams is that few of us have any real inkling how they work. Just what happens "under the hood" to all those jumbled-up data to make the graphs go up and down? What do non-specialists need to know about the inner workings of these models to be able to use the results in a sensible manner? How could a small organization working in the field, without any modeling experts to hand, tailor the models to address their own questions?

In the worst-case scenario, any method, from any palette of evaluation tools, can give the appearance of certainty where no certainty exists. We have all learned to be wary when the phrase, "Statistics prove that..." is used as a triumphant final shot to shut down debate.

The least worked-out of the dilemmas emerging during the La Jolla phase of this conversation was whether and how the models could be used without the computers. Because the computer-based models are a work in progress, because they are hard to comprehend, and because the people involved in this exchange are "can do" types, there was a tendency to explore other ways the content underlying the models could be put to immediate use. Would they serve as useful heuristics, a good framework for the trial-and-error of scaling impact, or checklists of priority factors?

Hesitancy

The La Jolla convening was filled with natural early adopters. Nonetheless, there was a certain amount of healthy skepticism. Several people knew from bitter experience the time that can be wasted building and validating frameworks, models, and tools that add little in practice or, worse, become

a distraction. So many times I have heard an exasperated voice say, "We don't need another logic model!"

So this is the question not only for these particular statistical models, but for any potential non-sequential evaluation tool: "What utility does it add?" The scale experts are already served well by a combination

At the same time, this middle ground feels uncomfortable. As Al Bartlett said, "Because the real world relationships we are trying to understand aren't always intuitive, you'd like to have the model. But the model isn't ready. So we fall back on the terms and relationships in the model – even though they are not validated."
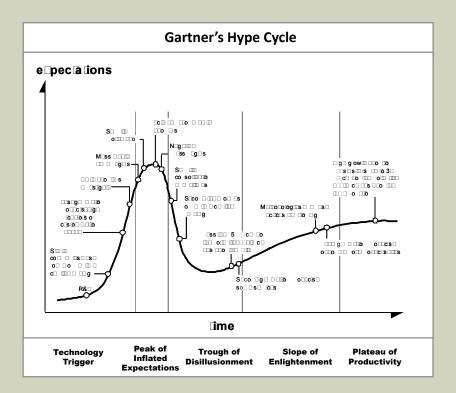
Potential

I have spoken of hesitancy, of limitation, and of dilemmas. Yet for all the uncertainty and doubt, La Jolla gave a warm reception to the work under review. The models embrace complexity. They depend on mathematics to function but draw data from all available sources, including local users. Their subject is adaptation and the models themselves can be adapted to reflect expert opinion and to take into account new learning. They are context-specific in the sense that they could be reshaped for Asia, or India, or Bihar, or any one of the Bihar villages that might benefit from an innovation seeking impact on family health at scale.

The models helped everyone in La Jolla to think differently about their work. They provoked a strong conversation about achieving impact in the face of both organized and disorganized complexity. They helped to visualize systems and they provided a benchmark against which existing tools could be measured.

The real potential of any method lies in its utility, accuracy, and simplicity of application. Tom Henrich brought us back to the GPS satellite navigation analogy. When it was first introduced, it provided a map, but no useful directions. Then when GPS began to offer directions, it was unreliable. "When it gave you a street to drive down you had to check that you wouldn't be going the wrong way down a one-way street. Gradually it has improved and you know it can be trusted. Which frees up your mind to do lots of other useful or interesting things. But if we had judged GPS on its first iteration, we would have abandoned it."

In addition to what a non-sequential evaluation method might deliver is the consideration of how it will help users to learn. In this respect, the models brought to La Jolla were a resounding success. The space created for the mutual identification of the factors that prime system readiness was universally welcomed. Engagement

Gartner's Hype Cycle

expectations

Technology Trigger | Peak of Inflated Expectations | Trough of Disillusionment | Slope of Enlightenment | Plateau of Productivity

time

The hype cycle is COIK

could see how Noshir Contractor was once a presenter on Indian television. He is an enthusiastic, exuberant, natural presenter, so much so that I was half expecting him to promote his model.

I could not have been more wrong. He was more articulate about its weaknesses than most and used a couple of illustrations that might be applied to any proposal for a new method.

For the first, he harked back to one of his professors at graduate school, who used to write "COIK" in the margins of Noshir's essays. Correct Only If Known. My co-authors scribbled it a few times in the margins of the draft of this synthesis, and it had the desired effect.

The second was an elegant graph drawn by analysts at IT research company Gartner Inc., called the "hype cycle." It describes a typical pattern of diffusion and dissemination for new technologies like the laptop, the smartphone and, dare I say it, non-sequential evaluation tools.

It starts with the "technology trigger" when the product is unveiled to its public. It passes through the "peak of inflated expectations" when there is a tendency towards unrealistic

enthusiasm, before entering the "trough of disillusionment" (I can still hear Noshir rolling the "r" in trough for added effect) when the failure to meet those unrealistic expectations leads to a loss of interest.

Fortunately, many innovations then climb the "slope of enlightenment" when the practical benefits become clear, before settling on the "plateau of productivity" during which market reach is established and the innovation evolves.

The models advanced by Noshir and Peter arguably went through all five phases of the hype cycle in La Jolla, and then came back to the "technology trigger," which is where we are left, uncomfortably poised.

Image: The Hype Cycle is one pattern of how innovations go to scale. (Adapted from Gartner's Hype Cycle Special Report for 2009, gartner.com)

Feedback loops

An electric sign reminding drivers of their speed will often slow them down.

This is feedback in the simplest sense, but it can be a useful art. Feedback *loops* are even more powerful. These are the virtuous and vicious cycles that, once started, can lead to runaway reactions, for good or ill. A rise in the earth's temperature is causing it to react in ways that make it hotter still. When shareholders hear of a fall in the stock market, they can find themselves rushing to sell, which drives down the price further.

The concept of feedback loops relies on the idea of a self-regulating system, a system that seeks equilibrium. The human body is a good example. Dieting to lose weight works less well the fifth time we try it than the first or second time. When we starve ourselves, our metabolic rate slows to help the rest of the body maintain equilibrium, a change that becomes fixed if we repeat the process many times over, lessening the impact of the diet.

Public systems have many of these qualities. Simple reporting of performance will alter performance. Sensing a change in equilibrium, such as the introduction of an innovation, the system will react and seek balance, until the point where the change becomes the norm. Each successive attempt to alter stasis can lessen the impact of future attempts.

The words "feedback loop" are becoming common currency in this conversation, so much so that we may overlook their meaning, or forget to try to measure them or to influence their impact on scale initiatives.

with the models demanded that participants worked through definitions, figuring out whether two terms meant the same or different things. It demanded disaggregation and re-aggregation of factors. It required that everyone expressed a view on the relationship between variables, whether A was related to B, whether that relationship was linear or non-linear, whether there were feedback loops, and more.

To my way of thinking this is a common language issue. How do we bring people together and get them to talk about things they might otherwise avoid? How do we replace what George Soros calls "momentary transactions" with sustained relationships among a diverse group of experts producing new forms of knowledge? How do we ensure that this talk leads to effective methods for scaling and evaluating impact? In short, how do we develop conversations such as this into a new form of scientific inquiry?

There was strong support in La Jolla for building a learning community around those participating in this conversation. There is strong momentum behind some of the emerging networks. But there is a need to further diversify the participants and to encourage competing and complementary networks, particularly if they can be rooted in Asian, African, or Meso-American contexts. Much of this potential is described in general terms, but I heard in La Jolla five clear proposals for practical application:

- To use the models to better understand the strengths and weaknesses of the many hundreds of initiatives sponsored by the US government in the preceding decades;

- To use the models to help policy makers in India to better appreciate and embrace the complexities of their aspirations;

- To use the models to reflect again on the nascent investments being made as part of the Salud Mesoamérica 2015 project;

- To establish new knowledge networks in Bihar, India, replicating there the space for experiment that has underpinned this synthesis; and

- To continue this exploration with another convening in India in 2012.

Conclusion

La Jolla was not conclusive, even for this particular thread of the larger conversation. It left many questions unanswered. We will return to some when we reconvene in India, and possibly in the discussions that take place along the way. Thinking about what might be done to advance the inquiry helps me to summarize the potential of La Jolla.

The READI framework helps to think about the influences on system readiness for scaling impact. The computer models help to make sense of the disorganized complexity that occurs when many factors collide, nullify, and multiply. They also exemplify some of the challenges that are associated with developing non-sequential evaluation tools.

As busy as the READI framework is, there are still missing sections. There is no mention of money. The macro-political environment is relatively neglected. There is a need to better describe the workings of informal as well as formal systems. Above all, there is a need for greater conceptual clarity about what is meant by "system readiness."

In La Jolla we brought methods to users. At the next convening we might ask users what they would want non-sequential evaluation methods to do, and how these methods would need to be packaged to be most useful. In other words, if we started with the questions that most trouble scale experts, and think about how they should be answered, would we arrive at a product that is similar or dissimilar to that pressure-tested in La Jolla?

The packaging part of this continuing conversation is important. With all the technology now available to us, there is a natural tendency to fight complexity with complex methods – and this is frequently a successful strategy. But there may also be scope for battling complexity with simplicity, to use a method as plain as structured conversation, convening, or the sharing of stories. ●

The web of life: connecting organized and disorganized complexity

Yesterday's complexity is today's clarity. What might be counted as organized complexity today was almost certainly disorganized complexity in the past.
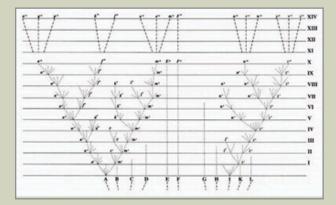
With advances in science and computing power, it is now possible to look at both the organized and disorganized aspects of certain phenomena.
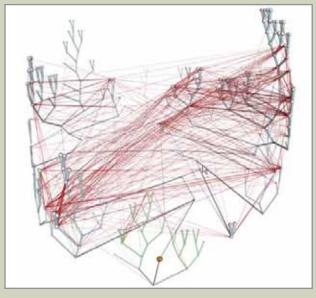
Darwin's tree of life, a simple diagrammatic representation of the origin of species that illustrates how genetic variation over time distinguishes between the species, is a masterclass in organizing what had previously been viewed as entropic. There can be few tasks more complex than explaining how the planet came to be inhabited by such a diverse group of organisms, but Darwin nails it in one diagram. This is some framework. In just a few sketchy lines, it captures a huge amount of organized complexity.

Darwin had no access to the ability of modern science to chart the types of bacteria that make up huge chunks of every organism. He might have dismissed this as disorganized complexity, important but dependent on too many variables to be knowable. But when networks of bacteria are overlaid on the tree of life, we suddenly begin to understand not only some of the potential causes of genetic variation but also the inter-connectedness of apparently disparate species.

Microsoft Bing designer Manuel Lima has termed this new representation the "web of life," and indicates some of the possibilities that might come with better mapping of the myriad of apparently random interactions between the components of systems upon which impact at scale depends.

Images: Top: Darwin's "tree of life." Bottom: the "web of life," showing similar types of bacteria present in different species. (Kunin, V. et al. 2005. "The net of life: reconstructing the microbial phylogenetic network." Genome Research 15(7): 954-9.)

Continue

Where will this thread of the conversation about achieving lasting impact at scale go next? Some of the next stages may move the location of learning to the Global South. But already it is impossible to foresee the many places the adaptive tensions generated in Seattle and La Jolla will eventually take us.

¿Hablas sistemas? I find it a little like my Spanish: I have moments of fluency when I both speak and think systems. But too often, because I am not a native speaker, this clarity is followed by moments where I am lost.

Native speakers of scale understand the personalities of systems of health organizations, delivery organizations, and the formal and informal components of government. They anticipate how these systems will adapt to stimuli, the extent to which their equilibrium can safely be disturbed, and the point at which a new equipoise will incorporate the innovation.

System speakers assess readiness like a clinician takes the blood pressure of a recovering patient, seeking a reading that falls below 120/80 but cognizant that there is no perfect number or figure, just a range that indicates a general state of health.

The people who speak this language value systems for their strengths – to systematize, to provide the infrastructure that will underpin the routine delivery of innovations in global family health, and to make impact an unquestioned part of day-to-day practice. They work to systematize pull, to get governments, health systems, and NGOs demanding what the user needs and wants.

There is not one language of system; there are several. The nouns, verbs and adjectives required to evaluate, package, and market a breakthrough in global health are different from those used to ensure long-term funding, get the professionals onside, or ensure that the right amount of product is in the right place at the right time. So systems speakers do a lot of translating, not only, say, interpreting for a health NGO the aspirations of an international donor, but also helping the catalyst to understand local constraints and solutions. They do more than translate words; they relay vital information about culture and context.

What they do is like what I hope to do when I broker the contributions to this conversation. They are using knowledge not as a product but as a practice, and their idea of knowledge is not restricted to scientific method; it extends to shared learning from the experience of trying to scale impact, with varying degrees of success, in different parts of the world.

Aptus, the fit

My role as knowledge broker has been made simple by the quality of the conversation. But important points of connection between potentially competing ideas have emerged from Seattle and La Jolla and will certainly underpin future exchanges.

The concept of "fit" is emerging as fundamental to scaling impact. In Seattle, the focus was on the link between innovation and user. In La Jolla, the focus shifted to the fit between the innovation and the systems in which the innovation will live. I have paraphrased the discussion by talking about the fit between the *catalytic system* (around the innovation) and the *delivery system* (around the user) – a shorthand for relationships among the many systems on which impact at scale depends.

In this part of the conversation, the join between scale experts' and scientists' mental maps has been explored, as well as those parts of complexity, organized and disorganized, that hinder that mapping. The result has been to propose a continuum between the traditional methods of science that are appropriate for sequential evaluation of a relatively ordered world, and the new palette of tools required to better understand the non-linear, dynamic, entropic domains. The call in Seattle for new methods should not be read as abandoning established approaches, nor as any loss of rigor.

Those who "speak systems" are using knowledge not as a product, but as a practice

As these discussions get physically closer to the places in which maternal and child health is most compromised, we will start to talk about the fit between the "global pull" and "local pull" for scaling impact. Already, there are the first signs that this conversation is itself devolving and re-generating in local knowledge networks in Asia, Africa, and Meso-America. The ideas that emerge in these contexts will challenge and test some things we are beginning to take for granted, leading to new innovations and solutions.

Tension

Adaptive tension – fuel for the fire of getting systems ready for scaling impact – will help to advance the Seattle-La Jolla conversation. La Jolla's "ghost," Trish Greenhalgh, cut to

the heart of two would-be orthodoxies when I shared with her the thinking from Seattle and La Jolla. "Context is king?" she asked, with more than a quizzical eye. "What is being lumped together under the banner of 'context'? There has hardly been," she continued, "any conversation so far about culture and community engagement, and there is far too little on the political economy dimension. After all, these major donors are doing a lot with money and influence." Unpacking context (and its regal qualities) might be a priority for future discussions.

Greenhalgh also made me reflect on the shape of diffusion of innovations. In Seattle, there were frequent references to Everett Rogers' "S-curve," the pattern of diffusion that looks like a smoothly rising escalator. But Greenhalgh asked me about the pattern of adoption called "Moore's chasm" – similar to Rogers' S-curve, but with a gap between the early adopters and the majority. She went on to talk about the "Van de Ven model," which is messy, non-linear, and incorporates shocks and setbacks: this seems to fit well with what I heard the experts say in La Jolla. Or maybe we need to look again at the "hype cycle" (page 31), where the sustainable level of expectation is eventually *lower* than the level in the early, heady rush of excitement?

And then it struck me. When we looked at the curves in the graphs produced by the La Jolla models, I didn't see any of these patterns.

As was suggested in the previous section, there is much ground still to cover in this conversation. Greenhalgh has helpfully indicated two of many areas where healthy tensions will exist among the ideas of those contributing. These tensions will be vital source material for discovering solutions to the problems that hinder impact at scale.

Action

Joe McCannon helpfully proposed that we try to use La Jolla to develop hypotheses that we could collectively test once we got back to our day jobs. For the most part, the objectives for the La Jolla convening were met, but on this count we failed. La Jolla will serve as a stepping stone to a longer, deeper examination of the issues in India.

That said, several participants left having made clear commitments to develop the

emerging ideas, to test the models against real-world examples of scale success and failure, and to re-examine proposed investment strategies in the light of the discussions. Noshir Contractor and Peter Hovmand both went away with data to improve their models, and the Kaiser team has a lot of currency to further improve their excellent summary of the existing scientific literature. Several hares are running.

In a separate publication, *How to Achieve Lasting Impact at Scale*, updated after each of the convenings, I will use the same concepts that underpin this synthesis to summarize the primary lessons for those involved in large efforts to improve human development.

At a more practical level still, I hope that beneath the broader connections sought by this synthesis, the reader can pull out some immediately helpful lessons about system readiness. Simple injections of innovation into large-scale systems are doomed to fail. Not taking account of power dynamics will handicap any initiative. True partnerships involve competition and strain as well as collaboration and synergy. Use people who speak systems and make the most of insider-outsiders at every level and every stage of the effort. Bundles of innovations may be more efficient to scale than a series of single innovations, given the amount of attention required from major donors and governments to make any given program a success. Demand and supply are equally necessary. Never forget that a system has history, and never imagine that the job is done.

If the view from La Jolla is "top-down," what happens when we look from the ground up? I was struck in La Jolla by the natural learning laboratories that are emerging in Bihar and Meso-America, or in the continued development of established programs like Saving Newborn Lives. These are the places that can accept Joe McCannon's invitation to collectively test jointly developed hypotheses.

I have said that knowledge is a practice, not a product; something we do, not something we consume. So how can we push the "doing of knowledge" into new arenas that would help scale impact? In La Jolla, there was a clear desire in the room for a continued and evolving learning community – one that will be open to newcomers who can make a contribution. Perhaps this learning

community, and others like it, could become places where the brokered knowledge applies not only to leveraging solutions (as we might call our current task), but also to managing implementation and becoming accountable for scaling impact in practice.

It is the emerging learning community, commenting on this synthesis as it is written, that also encourages me to reflect that we have been considering the innovation first, and the system second. What if we were to start with the system, and then think about what innovations it could sustain? To what fruitful enquiries would that lead us?

India and beyond

Where will this conversation go next? It is planned to reconvene in India late in 2012. But now that several hares have been set free, it is not possible to know where they will run.

The basic structure of the conversation is becoming clear. There is the fit between innovation and user, to which most of the Seattle meeting was devoted. There is the interaction between innovation and the systems that will nurture that innovation, the focus in La Jolla. And there are the tasks of putting the "L" back into the MLE of monitoring, learning and evaluation, and of finding scientific and non-scientific knowledge most appropriate for the choppy waters that impact-at-scale boats must navigate. I sense, as indicated above, that a healthy tension between the global conversations that have so far characterized this work and the more local discussions that it will sponsor could produce another productive seam to mine.

Between La Jolla and India there will be new knowledge. The mental mapping of the experts will evolve, as will the priorities of catalysts for change. But the potential to orchestrate new connections between emerging ideas will remain. The single most successful element of the La Jolla convening, engineered by Peter Hovmand and Noshir Contractor, was the way in which the computer models demanded that participants draw from their pool of shared knowledge to define and prioritize the strongest influences on system readiness for impact at scale. It is in this type of detailed, focused interaction among diverse groups of experts that the opportunities to leverage new solutions are greatest. ●

Participants

Alfred Bartlett is a pediatrician, epidemiologist, and director of Saving Newborn Lives, one of Save the Children's flagship activities to improve neonatal health and survival in developing countries. The program is supported by the Bill & Melinda Gates Foundation.

Noshir Contractor is the director of the Science of Networks in Communities (SONIC) Research Group at Northwestern University, where he is investigating the behavior of social and knowledge networks in contexts including communities of practice in business, translational science and engineering communities, public health networks, and virtual worlds.

James Dearing is a senior scientist at the Institute for Health Research at Kaiser Permanente Colorado, and co-director of Kaiser's Center for Health Dissemination and Implementation Research. Jim and his team created the READI report that was presented at the convening at La Jolla.

Christine Galavotti is the director for Sexual, Reproductive and Maternal Health at CARE USA, and is a technical advisor to the Bill & Melinda Gates Foundation's Avahan India AIDS Initiative.

Robert Hausmann is a managing principal at Hausmann International, a consultancy working on leadership and organizational change in complex organizations. He is completing his doctorate at George Washington University, focusing on social entrepreneurship.

Tom Henrich runs the Baby Care New Business Creation Team at Procter & Gamble. In his career in R&D, he led launches of a "preemie diaper" and a low-cost diaper in China and India.

Peter Hovmand is founding director of the Social System Design Lab in the Brown School of Social Work at Washington University in St. Louis, where he uses system dynamics to understand and evaluate community-level interventions.

Emma Margarita Iriarte is the principal coordinator of the Mesoamerican Health Initiative (Salud Mesoamérica 2015) managed by the Inter-American Development Bank (IDB) and sponsored by the Bill & Melinda Gates Foundation, the Instituto Carlos Slim para la Salud, and the Spanish Government.

Usha Kiran is the deputy director of the Global Health Program at the India Country Office of the Bill & Melinda Gates Foundation in New Delhi, where she leads the Foundation's state-based programming efforts in India.

Kendall Krause is a co-investigator for the Family Health Diffusion System Project, a project manager at the Institute for Health Research, and an evaluation consultant for the Division of Knowledge Management and Implementation Support at Kaiser Permanente Colorado.

Sarah Madrid is co-investigator for the Family Health Diffusion Systems Project, and is the program manager for the Cancer Communication Research Center, which is sponsored by NCI and directed by Jim Dearing.

Joseph McCannon is senior advisor to the Administrator and Group Director for Learning and Diffusion at the Centers for Medicare and Medicaid Services.

Wolfgang Munar is senior program officer for Family Health Solutions Integration in the Global Health Program at the Bill & Melinda Gates Foundation. He also works on Salud Mesoamérica 2015.

Brian Siems is a portfolio manager with the Bill & Melinda Gates Foundation's Family Health Division.

Rachel Smith is associate professor of communication arts and sciences at Pennsylvania State University. Her research focuses on the communication within and the structural patterns of social interactions, and their influence on health and wellness.

Neil Spicer is a lecturer in global health policy at the London School of Hygiene and Tropical Medicine and leads the qualitative component of the BMGF IDEAS (Informed Decisions for Actions) to Improve Maternal and Newborn Health grant.

Wendy Taylor is director of the Center for Accelerating Innovation and Impact at USAID.

Nana Twum-Danso has recently been appointed as a senior program officer for Community Health Solutions in the Family Health Division of the Bill & Melinda Gates Foundation. Previously, she was the executive director for African Operations at the Institute for Healthcare Improvement (IHI).

Rajani Ved is a physician and is advisor to the National Health Systems Resource Center for the Community Processes component of India's National Rural Health Mission.

Edward Wilson works for John Snow, Inc. as the IQC manager for the USAID Deliver Project, providing technical assistance for supply chain management and commodity security to USAID-supported countries.

Supporting the research teams

Erica Morse is a project manager at Kaiser Permanente's Institute for Health Research (IHR) in Denver, Colorado.

Franz Wohlgezogen is a doctoral student and lecturer at the Kellogg School of Management at Northwestern University, working with Noshir Contractor.

Mengxiao Zhu is a doctoral student in industrial engineering and management studies at Northwestern University, working with Noshir Contractor.

Facilitating and managing the convening

Thomas Backer is president of the nonprofit Human Interaction Research Institute, and an associate clinical professor of medical psychology at the UCLA School of Medicine. Tom co-facilitated the convening at La Jolla, along with Michael Little.

Polly Hogan is a program assistant in the Family Health team of the Global Development Program at the Bill & Melinda Gates Foundation.

Dwan Kaoukji is a researcher at the Social Research Unit at Dartington and a doctoral student at the London School of Economics.

Michael Little is co-director of the Social Research Unit at Dartington, and co-facilitated the convening at La Jolla, along with Tom Backer.

Beth Truesdale is an associate of the Social Research Unit at Dartington and a doctoral student at Harvard University.

Achieving Lasting Impact at Scale

Part Two: Assessing System Readiness for Delivery of Family Health Innovations at Scale

A convening hosted by the Bill & Melinda Gates Foundation
in La Jolla, California, March 29-30, 2012

Synthesis and summary by the Social Research Unit at Dartington, UK