Replicating Programs in Social Markets

David P. Racine

Replication & Program Strategies. Inc.

Revised August 10, 1998

Since the late 1970s, the federal government has been gradually shrinking its investments in social policy, looking to states, localities, charities, and even private businesses to bear more of the load and to make more of the choices. Among the questions raised but unanswered by the ebbing of federal leadership in social welfare is how, in the absence of an active or constant center, do the best approaches to improving social conditions get identified and their broader adoption fostered. Because of a common land, a common culture, and a common citizenship, American society retains an interest in extending the use of the most effective answers to social problems, even if the central government is playing a lesser role.

For most of the past half century, Washington was the principal place for innovators to take their promising new strategy or novel program in hopes of gaining wider notice and extending it to new venues. The aim was for Congress and the Executive Branch to help spread the idea wholesale around the country, using friendly policies and funding to induce lower levels of government to adopt and replicate it. Even if this hierarchical strategy failed to work as often or as well as conventional wisdom held, political circumstances at least made it a valid possibility. In recent years, however, that has become less and less true.

Today, suppliers of model programs and those in search of better ways to address social issues face something more like a marketplace than an intergovernmental hierarchy. Washington still sets some general parameters, but it is less involved in figuring out which social innovations work and in promoting and funding those that do. Suppliers and searchers are left more to their own devices in finding each other, much as they would be in a traditional market. The vertical integration of social policy — the hallmark of national responsibility for the poor and disadvantaged since the Great Depression — is being displaced by more numerous horizontal relations between the social-sector equivalents of sellers and buyers.

The question now is how to make this market-like arrangement function for the good of society and, particularly, for the good of those who need the help and support social investments can provide. A key answer lies in doing a better job of generating the information that markets in the social sector need in order to judge whether social programs are replicable.

What makes a social initiative a good bet for successful expansion? Replication is on old concept in social policy, but not one that has been well understood. The diffusion of promising programs through the vast and complex American republic has tested the wisdom and wherewithal of just about all who have tried. But as the social sector takes on more of a market dynamic, the factors that determine a model program's potential for growth are beginning to clarify.

Replicating the best, most market-worthy programs is, to be sure, not the only way to achieve broad social improvement. Advocating for better public policies and improved social conditions, strengthening the education and development of social sector practitioners, and enhancing the managerial capabilities of the sector's organizations all have essential roles to play. However, just as business markets need tangible products and services to sell to consumers, social markets need programs through which to apply what is known about improving lives and preventing harms. At the least, it seems unlikely that social conditions can change much without the broad replication of solid programs solving specific problems.

This paper presents a systematic method for weighing model programs and determining their potential for wider application. In particular, it sets forth a set of dimensions on which to measure the potential for replicating any given model, and describes how those dimensions can figure in a careful analysis of replicability. These ideas are based on intensive work my colleagues and I at Replication and Program Strategies, Inc., have done on more than a dozen replication efforts since 1994. These experiences have been supplemented by examinations of dozens more programs at various stages of replication.

An Imperfect Social Sector Market

On the surface, a more market-like social sector seems to serve the interests of democracy better. People at the state, local, and community levels have more freedom in deciding how to respond to social needs. But this freedom must be exercised under conditions that are strikingly more imperfect and complex than just about anything experienced in the traditional business sector. In general, business markets function imperfectly — that is, supply and demand fail to stay in equilibrium — because buyers and sellers lack adequate and timely information about all of their options. Information problems also hamper the performance of social markets, but in a more profound way.

The essential dilemma in business markets is lack of *access* to needed information. "I'm unaware that the car I want for the price I'm willing to pay is available at a dealership two towns away." The essential dilemma in social markets is not just that information is inaccessible, although it may well be, but that useful information is often lacking altogether. Decision makers can have all the information there is on a social program being offered to them and still not have what they need to make a reasoned choice.

This lack of essential information is less a problem in those areas where social sector consumers may directly exercise choice, as in health care or child care. There, consumer decisions (whether well informed or not) can contribute in a material way to the market's determination of the value a program or service acquires. But what about that large swath of the social sector where consumer choice is greatly restricted or nonexistent, as in many state-led programs to move welfare recipients into jobs? Or where direct consumers are only one voice among many influencing whether a program or service will be provided — as when a school decides whether to adopt a particular education reform strategy?

In these cases and throughout much of the sector, choices about what constitutes value and what does not are made politically. The sector may be getting more market-like, but its political character has not changed. And there is little reason to believe that it will, so long

as parties other than those who directly consume social programs and services are footing most or much of the bill. That condition, despite the diminution of the federal role, is likely to continue in the foreseeable future.

The strategic response to the social sector's valuation problem has been either to try to fill the information gap in an exacting but, from a market perspective, somewhat narrow way, or, conversely, to treat the issue as largely moot. Gap-filling has been the interest mainly of social scientists committed to the idea that a model program's effectiveness should be thoroughly evaluated before it is considered for broader adoption. The scientists acknowledge that evaluation cannot change the political nature of public investment choices. But they argue that it is still better for such choices to be informed by knowledge about what works and what doesn't than to be guided by political will alone.

Moving in a different direction, the moot response assumes that evaluation can shed only limited light on value. The large degree of "local content" needed for most social programs to work means that most evaluation findings cannot readily be generalized beyond their original places and circumstances. From this perspective of weak knowledge, it is normally better for communities to design their own models.

Both responses make telling points, but neither by itself constitutes a satisfactory guide to action. While conducting more and better evaluations is a worthy objective, the more rigorous kinds, such as randomized control group experiments, have proved expensive and often too slow to produce evidence when it would be politically most useful. Moreover, it is a rare evaluation that yields enough information to judge not only the effectiveness of a particular model, but what its market prospects are likely to be and how best to take advantage of those prospects. These latter considerations are almost routine in business sector investments once an innovative product has been determined to work.

The alternative approach — looking to communities for home-grown social program development — resonates with our nation's democratic spirit and tradition of localism. But

the belief that communities can routinely design their own successful models rests more on hope than evidence. The skill to craft programs that really work, like doing anything well, is not evenly distributed. Besides, it is far from clear that communities and states want to design all of their own initiatives. What they want, and what they have been getting from Washington, DC, is more authority to "choose" programs and services for their jurisdictions, based on the particular goals and outcomes they want or need to achieve. These choices may be of their own design or adopted from elsewhere.

Mitigating Market Imperfections

For social markets to work better, the response to the valuation problem needs to reflect the reality that a market for social programs does, in fact, exist. This acceptance means evaluative information must cover more of the factors that should drive decisions in a well-functioning market. Performance attributes, like effectiveness, are only one — albeit the most important one — of the factors that need to be considered. When states and communities entertain the possibility of adopting an extant program, they need information not only on whether the model worked when it was being tested, but whether it stands a good chance of working in their environments, how it will do so, and with what resource implications. Outside "investors" such as foundations and the federal government would benefit from similar information. And perhaps more than any of these, program developers themselves need to be more fully aware of what it means for their models to be choices in a social marketplace.

To be sure, these broader information requirements are not easily met in a sector that is only in the early stages of developing its identity as a setting for markets. But a useful place to begin is conceptualizing the factors that should shape decisions on whether to scale up a promising or proven model. What kinds of information do funders and others need for making good social investment judgments? There are no formulas for calculating the appropriateness of investing in the expansion of one program or another, as there often are

for weighing private sector investment choices. Still, from the experience with social program replication, it is possible to extract a provisional set of guidelines by which social investment decisions can take fuller account of the dimensions that will shape growth opportunities. These guidelines are defined by three fundamental realizations.

First is recognition of the necessary complexity of social programs and the environments in which they are designed to operate. Successful social endeavors are usually complex or uncertain because they try to change either the behavior of those they serve or of other organizations and institutions on which those they serve depend, or both. The uncertainty means that, in operating a program, adjustments constantly need to be made, based on experience, to improve the chances of favorable outcomes. Environments are complex because the survival of a social program depends not on profit and loss statements, but on an always-renegotiable political agreement among stakeholders. The combined complexity of these two sources means that the guidelines must themselves be fairly complex to be of real use in social investment decisions.

Second is the obvious but not always appreciated fact that a program is a solution to a problem. A problem has been identified in the environment (e.g., home-bound older people) and a program has been designed to address it (e.g., home health care services). To be a good solution, a program must match the requirements of the problem at which it is directed, and it must do so successfully and in a feasible manner. Accordingly, the guidelines need to deal both with the program's performance, that is, important qualities of the results it gets, and with the operations through which it performs.

Third is the importance of the fit between a program and its environment or context. A context will include both elements that are supportive of the program and those that are more constraining. Each must be considered in determining how good the program's chances are of being more broadly adopted.

Dimensions of Replicability

These general considerations enable the construction of a more specific scheme for identifying and arranging the dimensions along which a program will be found to vary when its potential for growth is being assessed. These are called the "dimensions of replicability" to indicate that an investor's interest in a model social initiative is not in mere growth, but in the model's ability, as it spreads, to keep producing the results that initially drew the investor's attention.

The dimensions can be plotted on a two-by-two diagram (see next page). The dimensions in the top half of the diagram pertain to the program itself. Those in the bottom half concern the "real world" context — the market — in which the program is to be applied. Dimensions on the left side relate to the quality of the program's outputs, to the social results it produces. How good are these "products" (upper left quadrant)? How supportive of their realization is the market in which the program is to function (lower left quadrant)? Dimensions on the right side of the diagram speak to how the program's outputs are produced. How is the program limited or challenged by the particular design of its production process, that is, by attributes of its working parts or elements (upper right quadrant)? How is this process affected by the market (lower right quadrant)?

Replicability can be assessed adequately only by considering all four quadrants — by looking carefully at what the program tries to achieve (i.e., the product) and how it tries to achieve it (i.e., the production process), and how both the product and the process will be shaped by the reactions of the market when the program is disseminated.

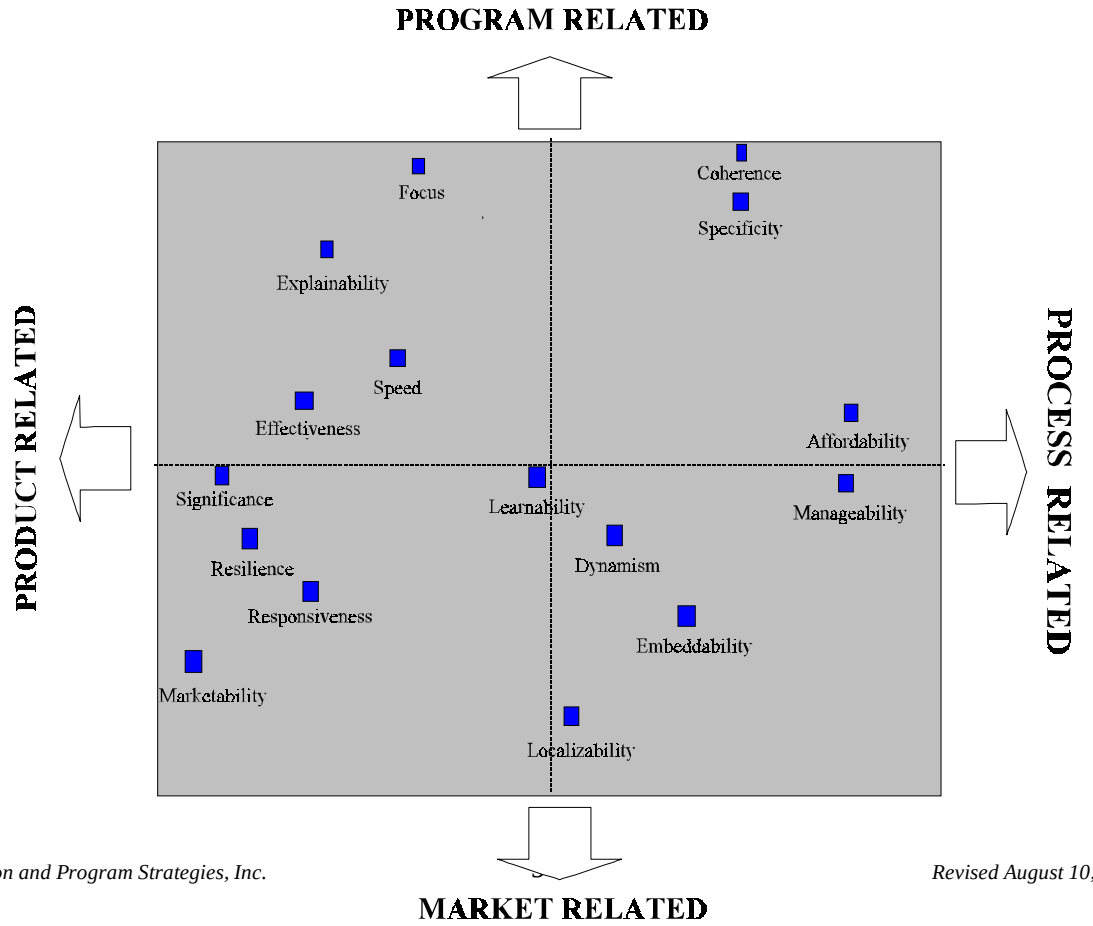
Each dimension has been placed on the diagram according to where our experience suggests it appears to fit best in relation both to program and market and to product and process. Consider, for example, the dimensions of effectiveness and coherence. Effectiveness is a quality of a program's outputs, which is why it shows up on the left side of the diagram. However, because in the social sector actual effectiveness is determined by both a program and its market or context, it is fairly close to the midpoint between the upper and lower

halves. By comparison, coherence is an attribute of how easy or hard the program will be to operate, which is why it lies on the process side of the diagram. It is far up on the top half because it is intrinsic to the design of the program, and for the most part unaffected by market factors.

It is unlikely that any program will do well on all dimensions. At the same time, a program that is weak on some dimensions may still be strong enough on others to have a good chance of expanding successfully. The point is that its prospects for growth will be affected by multiple factors that should be examined before investment decisions are made.

Program Product: How Good a Performer is the Program?

The first and most obvious set of dimensions speaks to key qualities of the outputs of a model program. Effectiveness is obviously crucial in assessing whether a program is replicable, but so are three closely related considerations: whether that effectiveness can be



explained, whether the program's aims are defined well enough to be achievable with increasing scale, and how quickly it gets its results.

In a perfect world, a well-defined standard or metric of ***effectiveness*** would help sort replicable social innovations from unreplicable ones. But in the imperfect world we have, standards of evidence in the social sector vary widely. They are unevenly applied, and they generally compare unfavorably to methods of measurement in the proprietary economy. There is no hard consensus about what the standards should be, and few institutionalized mechanisms are strong enough to force the issue. As a result, mediocre or untested social programs often make their way into replication because they have good marketing, or visible, charismatic leadership, or appear cheap relative to the alternatives. When such programs fail to produce the results promised, they contribute to disillusionment about social enterprise in general.

While a common evidentiary standard is probably beyond reach in the foreseeable future, the lesser standard of "requiring that a model be a demonstrable advance within its particular field in the social sector" would seem to be more approachable. The sector is divided into a variety of fields at different stages of development. Some fields — take health care services — have evolved rapidly over time and are governed, for the most part, by rigorous methods and standards for judging the efficacy of an innovative program or practice before it becomes a serious candidate for replication. Other fields, like youth volunteerism, have been slower to develop. They tend to be guided by more amorphous and variable definitions of effectiveness and techniques for measuring it. The substantive and methodological knowledge available for designing health care innovations is considerably more robust, at this point, than that available for youth community service. Because knowledge and standards vary, so should the role of evidence of program impact in analyzing replicability.

Yet evidence of positive impact by itself is of limited value. Statistical findings of effectiveness may appear impressive, but unless they can be ***explained*** as having been caused by the program as designed, rather than by other, unknown or unacknowledged factors, there will be reason to question how true the measure of performance is. Admittedly, causation is often difficult to ascertain when, as with social programs, the intent is to change the way willful people and organizations behave. Nonetheless, when investing limited resources in the expansion of social endeavors, the advantage should lie with those innovations that can make the most plausible case for a direct connection between their planned activities and the anticipated results.

Ideally, programs are able to draw upon sound theory to demonstrate the relationship between what they do and the outcomes they produce. Take the well-regarded Prenatal and Early Childhood Nurse Home Visitation model, a program that helps low-income women give birth to healthy babies and enables mother and child to get off to a good start as a family. After several years of testing and development by researchers at Cornell University and the Universities of Colorado and Rochester, this program is just now being more broadly disseminated. It has been designed on the basis of established theories about adult motivation (self-efficacy theory), parent-child relationships (attachment theory), and environmental influences on family life (community ecology theory). The highly positive outcomes the model has achieved during testing can be explained relatively well using these theories. Such causal associations do more than provide potential investors with comfort: They make the program more comprehensible for those who will implement it in new settings.

Of course, not all promising programs will be grounded by their designers in known theories about human development and change. At a minimum, though, they should be explainable logically. They should evince a plausible argument for how the elements of the program lead to the results that make the program attractive. No matter how good the results

appear numerically, if the program cannot be causally explained in this minimal way, its replication will probably not be worth the investment.

Because of the intrinsic looseness in the standards by which the performance of social programs can be judged, evidence of impact and its explainability rarely reach a level that identifies a model as a sure bet for successful expansion. There is always room for doubt about how a program will fare as it is scaled up. This is why it is necessary to gauge how well the intended outputs of the program match the requirements of the problem it is supposed to address. A program that has been designed to perform in a ***well-focused*** way, representing a good fit with its problem, will be easier to keep focused as it spreads than one that aims more broadly. For example, a program that seeks to "increase the time parents spend reading to their school-age children" will, in all likelihood, be easier to hold on track than one that purports to "help families function better in a changing society."

The latter type of program, because of the breadth of the problem at which it is directed (i.e., family functioning), may attract a wider audience. But it also is liable to result in greater variation in what the program looks like and does from one place to the next. Some variation across localities, of course, helps a program remain fresh and evolving, not to mention responsive to differing local conditions. But too much variation may lead to local efforts with so little in common that they cannot reproduce the same or similar outcomes reliably, or even communicate effectively with one another. Arguably, it is easier to extend a well-focused program and enable it to change with time and circumstances than to start with a less carefully targeted one and try to bring it into sharper focus, and thus stronger impact, as it spreads.

The problem of focus is significant in the social sector. This is where society's highest ideals about human development and moral regard spur much of the action. Aspirations run high, as programs or program-like efforts seek to end poverty, reform public education,

eliminate child abuse, or overcome one of the other dilemmas that plague people in modern, changing societies.

While the social sector's idealism is one of its most important assets, enabling it to perform a necessary function as society's conscience, such far-reaching hopes eventually must be translated into practical measures that have staying power and can be built upon. A willingness to "work up and out" from carefully focused programs that have been or are being replicated may represent a credible alternative to the more sweeping reformism that has often punctuated the sector's development and captivated its leaders.

A final dimension of performance when considering replicability is the ***speed*** with which a program acts to generate worthwhile outcomes. Though social problems are complex and do not succumb to quick fixes, there is no getting around the reality that we live in a culture that tends to favor fast action over slow. Public and private funders of social programs typically want to see positive, demonstrable results within one to three years of making their investments. There is no scientific or rational basis for imposing such time frames. But they afford an efficient, conventional way for funders to allocate limited capital, perhaps in the hope that at least some investments made under these arbitrary circumstances will pay off.

The lesson in this for program designers is not that they would be better off pursuing short-term strategies only. Many do, and they end up with programs whose effects are too negligible to be worth sustained investment. The challenge is to devise models that, while they may take relatively long to achieve their most significant outcomes (e.g. delinquent youths' entry into gainful employment), can meanwhile yield favorable results along the way (e.g., reduction of delinquent youths' participation in criminal activity). Interim performance indicators are important both to funders, whose money is at stake, and to program staff, who derive encouragement from signs that they are making headway. The more replicable models will be those that take the need for speed into account without completely capitulating to it.

Program Process: How Easy is the Program to Implement? Often a program's performance is the only standard by which the wider world judges it. But to gauge a program's replicability, other dimensions related to the ease of implementation may be equally important. Whether the program's components are specific enough to be identifiable and understood, and how well they fit together, will influence whether it is easy or difficult to operate. There is also the obvious constraint of direct costs — whether a particular operational design is affordable, relative to the benefits it produces or by other standards.

Specificity in program design is relevant for the simple reason that it makes the program's components easier to understand and adopt. Research on adult development indicates that mature people tend to learn mainly by analogy, that is, by comparing new information with what they already know from experience. In the social sector, such experience typically involves the design and delivery of particular kinds of services. As a result, the usual social sector practitioner is likely to understand and respond more quickly to a new program that is well-described operationally. For example, in studying education reform designs disseminated under the auspices of the New American Schools Corporation, researchers at the Rand Corporation found that early replication was most successful for models that, among other things, were spelled out more concretely and completely. Teachers and administrators evidently respond best to innovations that resonate with their experience as practitioners and that do not require too much extra cognitive work to appreciate.

Programs that lack clear and specific components are not necessarily unreplicable. Under the right circumstances of inspirational leadership, high commitment to a popular cause, or strong demands for social change, a substantial portion of the audience for a program defined conceptually may be convinced to adopt it despite its vagueness. The chances are good, however, that significant problems may arise subsequently.

Just as with a program with broad aims, one that is operationally underspecified may draw adopters who are not really capable of implementing it successfully. Because the

program has not been described in adequate operational terms, they buy into an attractive idea or set of principles only to discover, when trying to operationalize it, that making the idea or principles work is beyond their ken. To be sure, sometimes the very purpose of offering a loosely specified program for replication is to encourage adopters to invest more of their own creative energies in applying it to their different environments. Yet replication efforts that begin by granting wide latitude in defining program operations will often have a hard time achieving more consistency among adopters later on. Discretion, once granted, is rarely surrendered for the sake of standardization, even if that seems necessary to success or even survival. It may be easier for a well-specified program to "discover" — over time and with careful testing — a wider zone of discretion for its local operators, than for a more broadly defined program to evolve toward more shared specificity.

While the components of a program may be carefully described, it is another matter for them to ***cohere*** as a whole. Do the parts reinforce and complement one another, or have they been patched together without sufficient regard for fit? A coherent program is not necessarily one in which the components are tightly bound up with one another, such that changing one part deeply affects the operation of all the other parts. Rather, coherence means that the interrelationships among program elements, whether tight or loose, are clear and understood and informed by the same assumptions or theories about what the program is trying to accomplish. When a component changes in a coherent program, the effect of that change is more readily anticipated than in a program lacking coherence.

The home visitation program discussed earlier offers a good example of programmatic coherence. The program focuses on poor women, usually teens, who are experiencing their first pregnancy. The main reason for this focus is that researchers who designed the model believe that nurse home visitors can establish rapport sooner and more effectively with women who, because this is their initial encounter with pregnancy and motherhood, are more open to information and guidance. In other words, the program has aligned in a fairly precise

way the definition of its target group with the type of relationship visitors need to establish to be effective in their work. This same kind of internal attunement has been pursued in developing the other elements of the program: detailed visit protocols organized around key areas of family and individual functioning, a visit schedule varying in frequency to reflect different stages of the mother's and her child's development, involvement of other family members and friends, connections to other community services that participating families need, and specific qualifications for the nurses who serve as visitors. Aligning all of these elements with each other has been no simple task. It has taken several years to test the model and fine-tune it to the point where it is robust enough to be considered for broader replication.

Programs with limited coherence can and do get replicated. In these cases, the burden for making the program more coherent rests mainly with its local adopters. It is up to them to fill in the gaps left by incomplete linkages in original program design. In doing so, they are apt to be heavily influenced by the need to make the program cohere on the basis of local understandings of success or fitness. Those understandings may turn out to work, or they may not. But in either case the result over time is likely to be a program in which the crucial features differ from one place to the next. By contrast, when the starting point is a coherent model, divergence in local operations will still occur as time goes by, but it will be around a core set of well-defined, integrated elements to which all operators subscribe and through which they can comprehend and learn from one another's experience.

The operational requirements of a program are what drive its costs, and these costs in turn influence whether or not it will be perceived as affordable. ***Affordability*** is not an absolute or even mainly an objective criterion, but a function of people's perceptions of the relative value or advantage of the program within the environment or context in which it will operate. It is a matter of comparison, of how a given program compares to what already exists, or with what people already are, or think they are, doing or can do.

Affordability can, and perhaps should, involve a deliberate calculus. Yet uncertainties about social investment in general often induce people and organizations to fall back on simplifying, unreflective biases about what a program ought to cost. Potential adopters and investors may make judgments of affordability on the basis of several kinds of comparisons.

Almost always, the hope is that a program's expected benefits or outcomes will exceed the expected costs. Unfortunately, quantifiable evidence on the benefits and costs of social programs is hard to come by. More specifically, while hard data on costs are usually available, information on anticipated benefits is rarely rendered in the same, monetized way. For example, we may know how an at risk youth is likely to benefit socially from being involved with a Big Brother or Sister (e.g., improved participation in school, less getting into trouble), but cannot readily express this gain in economic terms. Because of the difficulties of translating benefits into financial gains or savings, other, more subjective comparisons become necessary.

An obvious basis of comparison is with the known alternatives to a program. Alternatives include both similar innovations (i.e., similar types of service focused on solving the same problem) and the related programs and activities in which potential adopters and funders are already invested. In the absence of hard cost-benefit evidence, these comparisons are often driven by differences or similarities in philosophy — weighing, in a sense, what a program represents against what members of its audience or market hold to be true or right. The "differences" side of the ledger will tend to be enunciated by those already aligned or sympathetic with an alternative or rival model, who may have reason to exaggerate the program's allegedly negative or questionable features. In this form of comparison, the issue is not so much the program's affordability in dollars and cents as its value in philosophical terms, its fit with principles held by various segments of the intended market of possible adopters.

Consider the nurse home-visitation program mentioned earlier. It is often compared to another home visitation model that uses paraprofessionals rather than nurses. While a paraprofessional version can be expected to be cheaper, the comparisons made between the two versions have focused mainly on whether there is an advantage in the visitors' being more like the women they are trying to help, as is the case with paraprofessionals. There is in fact little evidence that such an advantage exists. But a belief that it does, can, or should exist persists among many of those who have adopted or are sympathetic to the paraprofessional approach. Or take Big Brothers Big Sisters. Although it is the only youth mentoring program for which there is persuasive evidence of effectiveness, it is often compared unfavorably to less structured forms of mentoring that are more lenient in defining who may qualify as a mentor.

While comparing programs with their alternatives is probably the most common approach, another is to compare them with costs and practices considered customary in their field. By that reckoning, a program would be judged more affordable to the extent that its costs are in line with (or lower than) generally accepted costs for similar services. A related basis of comparison is the spending tradition of a given place or organization considering adopting a program. One state legislature or city council or philanthropy may be comfortable with the program's costs because it has spent that kind of money before, presumably, without adverse consequences, while another balks because the cost falls too far outside its norm or tolerance level.

In today's social markets, the affordability of a program becomes increasingly central in judging its replicability. Consequently, what is cheaper and modestly effective may often win out over what is more effective but costlier. It is, after all, almost certain that the money for a new program will be spent if made available; less certain is that the spending will produce the expected social results. Proponents of relatively expensive programs must be

able to make a compelling case for the ability of their models to deliver strong results if the investment were to be made in their growth.

Market Supports: How Encouraging is the Environment for What the Program Tries to Achieve? A program's broader market or environment will be more or less favorable to the kind and level of social change it claims to engender. The people in that environment will have their own, no doubt varying, views about the program's social significance. More specifically, those who would have a direct stake as operators or supporters will assess how well the program suits their particular interests. And the likely operators will determine how easy or hard the program will be to learn, based on the competence they bring with them and the complexity of the model relative to that competence. The market will also be a key source of feedback about the performance and impact of the program as it progresses. The kind, frequency, and amount of feedback the program attracts will influence its ability to stay focused and its responsiveness to relevant environmental conditions.

In one basic sense, a program in the social sector is like any product or service in the regular for-profit sector. To have a legitimate shot at growth, it must appeal to a large enough segment of the market. Much of the appeal of a social program turns on how a wide variety of actors — potential adopters and funders, opinion leaders, social scientists, and others — perceive its ***significance***. Does the program address a salient problem in a non-trivial way? While this appears to be a straightforward question, it is complicated by the fact that significance is a relative criterion, and thus easily disputed.

Many people will agree on what social problems are important. They may even define problems the same way. Where they will frequently differ is in their beliefs about the most appropriate solutions. What one ideology or political position sees as a significant and worthy solution another finds wrongheaded. Since all programs reflect particular values, no program is completely beyond becoming a disputant in ideological debate.

Some programs are designed to express visibly, even aggressively, a particular set of values, often in opposition to those who hold other values (e.g., liberal vs. conservative). Such programs, while they may serve the valid purpose of advancing social debate and illuminating important issues, are usually not good candidates for sustained growth. Their strong values prompt those holding other values to weigh in against them. The result is to stymie their market potential.

Better candidates, the ones most likely to be deemed broadly significant, will either be programs that rest on values to which many or most people can subscribe, or those whose public image is so attractive that their value orientation is muted. A good example of the former is Habitat for Humanity, the organization that uses volunteers to build homes for low-income people. Its religiously informed ethic of volunteer service applied to the practical task of home-building has enjoyed wide resonance across the political spectrum. In the latter category are, for instance, programs developed principally through scientific means, such as the already noted home visitation program. Models rigorously designed on the basis of established theory and knowledge and carefully tested, while hardly value-neutral, may be able to make the evidence of their effectiveness the primary quality by which they are known. They lead more with their impact than their value assumptions.

Social significance is a general, somewhat ambiguous judgment. Even if a substantial set of actors in the marketplace regard a program as significant, that may not be enough to prod the market into broadly adopting and diffusing it. Most of the time, a program must also meet the more concrete test of ***marketability***. That is, it must appear to respond in a direct way to the specific needs and interests of those who will operate and fund it, and to the policy mandates and goals under which these operators function. Consider the example of the federal welfare reform law enacted in 1996. The new policy demands that states move welfare recipients into jobs sooner rather than later. Programs that address this need for quick movement off the welfare rolls are likely to fare better in the marketplace of state decision-

makers (i.e., gain broader adoption) than programs that focus on developing recipients' longer-term employability. The latter may be viewed as significant by many people, but it is not what the buyer's market, made up mostly of state officials, is looking for just now.

Most social programs are not initially designed with a market in mind. They are created locally or for research purposes, with little or no attention to the people and organizations to whom they may eventually be promoted. Notice the contrast here with the proprietary economy, where trying to grasp what the market will buy often comes before, not after, product development. This tendency of social programs to ignore or downplay market considerations during their development leaves their designers with an unenviable task later, when replication becomes a possibility. They find themselves scrambling after-the-fact to understand the proclivities of their likely audiences.

This is more than just a matter of how to position a program or how to depict it most attractively for the marketplace. There is a strong possibility that, once it becomes clear what the market wants, the program itself may have to change if it is to have a decent chance of spreading. For designers heavily invested in what they have already designed, this kind of belated discovery is hardly ever welcome. Still, an ability to read the market accurately and respond accordingly appears to be as important in the growth of social programs as it is in the spread of commercial products and services.

While marketability is highly desirable, indeed necessary most of the time, it can also be overplayed. In a sector where standards of effectiveness are underdeveloped, the market of potential adopters will often, as we have seen, rely on other modes of assessment to decide whether a given program is likely to supply their needs. They may, for example, be attracted to models that are perceived as different or innovative, or to designers who are perceived as significant in their own right, apart from the program they have created. Some may simply follow what others in their network of contacts are doing. As a result, programs may spread

because they appeal to the marketplace but not because they have actually been shown to work particularly well.

This occurs in the for-profit world, too. The difference is that there is a greater chance inferior commercial products will be found out and taken off the market or changed as consumers express their displeasure or gravitate to superior products. In the social sector a mediocre program can endure for much longer. Because its effectiveness is hard to establish to begin with, firm judgments about its ongoing performance are difficult to make. The program need not perform measurably well to survive, and perhaps even thrive, so long as it satisfies other valued needs, such as linking its operators into a network with political clout or having their status enhanced by being affiliated with a highly reputable founding organization or person. While these alternative values are not entirely to be gainsaid, they are hardly acceptable as a key basis on which investments in social initiatives should be made. Being marketable is a good thing so long as the program being marketed has a strong likelihood of fulfilling the social purpose for which it has been designed.

Whether a program can fulfill its purpose will also depend, in no small way, on how ***learnable*** it is. Generally, a program is not considered for replication unless it is innovative — that is, unless some significant group of people see it as an enticing departure from current practice. But because it is innovative, it requires those who operate it to acquire at least some new competencies. The issue for replicability is not that some learning is necessary, but that the amount of required learning will affect how people will respond to the prospect of replication. If the program requires prospective operators to go well beyond what they already know and are skilled at doing, they may have a hard time developing the confidence to implement it successfully. On the other hand, if the learning necessary to operate the program challenges them too little or not at all, they are apt to see little reason to try. The trick, obviously, is finding the right balance, which demands a thorough understanding of the operating competencies the program entails and a firm grasp of the

extent to which potential adopters have an adequate foundation for developing these competencies.

Getting at a program's "learning requirements" can be made harder by the fact that pilots or demonstrations do not always reveal what it will take to run a program under the less controlled circumstances of replication. Typically, the pilot yields information about how to operate the program in and of itself, but little or nothing about how to negotiate for the resources, cooperation, and legitimacy it will need from the environment where it will be implemented next. Success on the latter score is often mistakenly assumed to come from general political talents or from the ambiguous ideal of "strong leadership," rather than from specific talents called for by the program itself. But success in getting an innovative program going well in, say, education demands different environmental know-how from that required in social services or health care. These different contexts vary in institutionally relevant ways that need to be taken into account. Replicability will be greatest when the designers make an effort during initial development to comprehend, in terms of necessary competencies, both the operating requirements of the program and the negotiating requirements of its particular kind of environment. While there will almost always be much more to learn than can be captured in any model, that is no excuse for neglecting context as a learnable element when a model is being developed.

Most social programs must function today in somewhat agitated markets. The social sector has become more turbulent as tighter constraints on resources intensify the competition for them, and as the authority to allocate resources grows more diffuse. Turbulence not only makes learning from the environment harder. It also complicates the feedback on which programs must rely to gauge their performance.

A market sends a continuing stream of signals about how a program is doing. In addition to information on how well the program is meeting its particular performance objectives, much of which will or should be captured by its management information system,

the market generates data on political support, resource possibilities, competing and affiliated programs, public opinion, and many other factors. While some of the signals sent are relevant, not all are. Programs always face the challenge of separating the relevant from the irrelevant feedback, much in the same way that a business must decide which market signals to pay attention to. ***Resilience*** is the ability to do this successfully, such that a program's performance is constantly benefitting from the environmental data that are absorbed and used by its operators. Ideally, little or no time and resources are wasted processing and acting on extraneous or noisy data.

When markets become turbulent, however, distinguishing between good and bad information gets harder because there is more noise to contend with. The chances increase that a program will act on feedback it would have been better off ignoring or will ignore feedback it would have been better off taking seriously. Programs able to minimize these "false readings" will be the stronger candidates for replication.

To be sure, the ability to read a program's market for feedback relevant to its performance is mainly a function of the competence of its operators. But it is also affected by the orientation to performance built into the design of the program itself. Well focused programs with clearly defined outcomes, even in turbulent environments, generally are more able to pick out the pertinent signals through a more intensive form of attention than is possible with programs that are more broadly aimed. In the latter case, breadth in what a program is trying to accomplish leads to breadth and variety in feedback, increasing the information processing burdens on operators and the likelihood of their making "false readings."

Consider Family Matters, a program developed by the Points of Light Foundation to engage families in volunteer community service. During its pilot phase, Family Matters was designed broadly with the aim of increasing the numbers of family volunteers. The six local sites participating in the pilot had to decide for themselves what the program would look like

(e.g., what kinds of families to recruit, the types of volunteer service they would do, etc.) based on their readings of local circumstances. A couple of the sites misread their local environments and as a result ended up with programs that were not viable. Subsequently, based on this experience, Family Matters has become more sharply defined and focused, and is now being more widely replicated.

This is not an argument against broad program designs *per se*. It simply points out how much more a broadly defined program must rely on its operators' talent for handling large and varied amounts of data from the environment. It takes a high level of skill to interpret the market accurately when the design of a program offers little or limited guidance.

At the same time that replicable programs need resilience to preserve their integrity in a dynamic context, their performance also needs to remain ***responsive*** in one degree or another to that very context. With resilience, the challenge is to distinguish between useful and unuseful information. With responsiveness, the issue becomes the quality and timeliness of the "useful" information. Does a program's market produce feedback that is clear and specific enough, and at the appropriate junctures, to judge performance accurately and adjust operations if performance is not meeting expectations? For example, a welfare reform program that trains recipients for employment needs accurate and timely information on job availability in the local labor market. If it cannot get such information, it may wind up training people for jobs that do not exist.

Programs do not all have the same dependency on feedback from the contexts in which they operate. In some fields, programs may benefit from a sort of institutional fence that limits their need to respond to environmental signals. For many years this was the case, for instance, with public education. New models could be introduced into schools or classrooms without much concern about how students' parents or the larger community might view such changes. This is less true today after the "loosening" instilled by two decades' worth of education reform, although the educational fence remains stronger and

higher than it is in many other fields in the social sector. The same institutional protection existed in health care, until the combination of rising cost worries and consumer concerns about quality forced that system to become more open to environmental influences. In today's more turbulent social sector, such forms of institutional and professional sanctity are becoming less common. And the importance of responsiveness has grown in proportion.

Market Constraints: What Limitations Does the Environment Impose on the Operation of the Program? Besides generating feedback of direct relevance to a program's performance, the market is a source of constraints that must be considered in trying to understand how the program will need to operate. Under ideal circumstances, a program would simply be able to operate as it was designed. In reality, however, its manageability will depend on how complex it is to operate given its context. There will be other questions, too: How much variation in local settings can it accommodate without compromising its essential features? How easily will it mesh in these settings with the other programs and systems on which it may depend? And how well will it adapt over time either to improve or maintain its performance?

A program's ***manageability*** is essentially the operational counterpart of its learnability. The ability of operators to learn the program will, in most cases, be strongly correlated with their ability to manage the program once it has been learned. The correlation is not perfect, though, which is why manageability merits being seen as a separate dimension of replicability. It is not perfect for a couple of reasons.

One is the obvious fact that the abilities involved in learning and those involved in managing, while they overlap, are not identical. Some people may be avid learners but, compared to those whose learning is less efficient, they may turn out to have more difficulty in making a program work day-to-day. The more halting learners may be better at handling interpersonal relationships or negotiating conflicts or executing routines that are integral to running most kinds of social programs. Of course, many people are talented in both ways,

but this is something that, rather than assumed, needs to be established by knowing the audience of potential adopters.

The other reason why learning and managing aren't perfectly correlated is that programs are inevitably more complex in their actual operations than any kind or amount of focused learning can encompass. It is commonplace to hear people involved in replicating a model talk about the differences between what they learned when being trained to run the program and what they later experienced doing the work. Obviously, the more reality training can reflect, the better. But there is always a gap between operating knowledge that can be articulated and that which remains tacit or arises from the seemingly endless variety of new circumstances that training cannot anticipate. The greater the gap, the more "artistry" operators have to exercise in managing the program. Investors need to consider this gap in determining whether a model program, which may score well on many other counts, is something that can actually be operated successfully under the more varied and less controllable conditions expected with scaling up. The more artistry required, the more risk.

The manageability of a program also brings up the important distinction between adopting a model and assimilating it. People and organizations adopt new programs for all of the reasons that have been previously cited. They may be attracted by the evidence of effectiveness, the compelling quality of the ideas or principles reflected in the model, the charisma or reputation of the developer, and so on. Yet once the program is adopted, they begin confronting specific realities that beforehand they could only have imagined generally. They must make choices about operationalizing elements that they previously perceived more abstractly. The burden of these choices lies with them, and thus so do the attendant consequences. As a result, their inclination may be to hedge, to opt for small steps that can be reversed if conditions seem to warrant. The assimilation of the program thus proceeds slowly and sometimes stops before getting very far. The clearer a program is up front about what it will probably take to implement and manage it well, the easier it will be for

prospective adopters to decide if the program is right for them. This clarity will not assure effective assimilation by those who opt to adopt, but it should improve the odds.

A key constraint in trying to spread a program is its ***localizablity***. Under traditional notions of replication, the assumption has been that replicable models should be those that can work just about anywhere. This bias has informed much social science research aimed at designing effective programs over the past thirty years, and it remains the orientation of most public policy-making in the social sector. While good programs of universal applicability are highly desirable, they are also extraordinarily rare. They are rare for the simple reason that there is too much variation across local settings in a diverse country like the United States. It is almost impossible to come up with strong models that can perform well under the different conditions represented by all of these settings. Consequently, it makes little sense to try to force a program into an environment unsuited for it.

That said, it is also true that the most replicable programs will be those that can work well in the greatest variety of environments. This means that their essential features — the elements that allow them to generate good outputs — can be preserved as they gain adoption in new places. But by implication it also means that some aspects of their design can be modified to accommodate local idiosyncracies without harming performance. While differentiating between the essential and the adaptive is not always easy, it is a crucial part of the calculus that goes into determining a program's replication prospects. Programs that beg this question create more ambiguity for their potential adopters, increasing the likelihood that some will adopt who should not, and others will not who should.

Just as no person is an island, no program works alone. Every program relies not only on the people tasked with operating it, but also on the cooperation or acquiescence of other organizations and people in its environment. A home visiting program for low-income mothers depends on the availability of other health and social services that these mothers and their young children may need. A training program for welfare recipients depends on

connections to employers who can provide jobs to its graduates. In this sense, a program being replicated in a community needs not just a host organization; it also needs a host context. This is what is meant by ***embeddability***. It refers to how well a program will integrate into the local settings where it is adopted.

Programs often get replicated but fail to take root. On the surface, the reasons may appear obvious: funding ran out, leadership changed, priorities shifted. But underneath the obvious explanations often lies a more fundamental failure to embed. All of the people who needed to be committed to the program in a given local setting were not, thus making its erosion or elimination easier.

This is seen with some frequency, for example, in school change. Efforts to use the schools as settings for the implementation of other social programs have met, at best, with mixed success. Such programs may get adopted but then are dropped when startup funding ends or the initial enthusiasm that greeted them dissipates. Ostensibly, the explanation is that school faculty have a hard time becoming strongly committed to efforts that fall outside the core curriculum. While this may be true, teaching practice is not immutable. It is decided by people — teachers, administrators, parents, state policy makers, professional associations. That is, it is a product of a set of relationships. If changes in teaching practice are to occur, then the relevant network must go along or at least not care, or else a rival set of relationships with sufficient power must be recruited to the cause.

Any new program, especially one being imported from outside, disrupts existing practice. It conflicts with some current interests and threatens some current relationships. To become successfully embedded, a program must be able to overcome or at least "wait out" these potential sources of resistance both within and outside the operating organization. Models that are clear and specific about the traps that need to be run for their embedding to occur will generally fare better in replication than those that are not clear and specific. This means that while a program is under development, attention needs to be paid to the kinds of

environmental connections and relationships on which it is apt to depend when replicated. Indeed, the most central of these linkages should be a feature of the model itself, so that they can be taken into account when pursuing sites for the program.

A program's embeddability is also often a function of time. The longer something is around, the more likely others are to relate to it positively. More time generates more knowledge, and more knowledge is the basis of a more lasting form of trust. Every social program either defines explicitly or entails suppositions about the length of time a person, group, or community needs to participate in order to benefit from it to the extent expected. Typically, adopters are willing to let a program go through one complete cycle before deciding whether to continue or discontinue it. A program with a short cycle, say of a few months, may be attractive because of its fast pace and associated low cost, but it also can be dropped sooner and more readily. Adopters will have had less time to get to know the program and to develop their commitment to it. By contrast, when the cycle is longer, the commitment of those who eventually adopt it may be stronger, even though the decision to adopt may be reached more reluctantly because of the costs involved.

It is important to distinguish embeddability from institutionalization. People and organizations with programs to replicate often see institutionalization as the ultimate aim of their efforts. While it is not wrong to hope for institutionalization, neither scholars nor practitioners have figured out how to strategize toward that end. To institutionalize a program means to incorporate it so deeply as the way to address a given problem or need that its continuation can be essentially taken for granted. Yet, institutions are seldom the result of deliberate effort. They are more the accretive outcome of a long, meandering, and largely uncontrollable process that eventually settles on a particular form of activity. Embedding, by comparison, is a more practical and accessible process. It involves the specific work undertaken to fit a program into its setting — the moves made to gain cooperation and

comfort from other influential parties. In time, institutionalization may occur, but there is no guarantee, nor an especially compelling reason to expect that it will.

This brings us logically to the final dimension of replicability: a program's ***dynamism***. A dynamic program is one that is able, in the settings where it has been adopted, to remain useful and possibly even become stronger over time by adapting, as necessary, to relevant changes around it. In other words, it can be improved and sustained in response to its changing market. Clearly, contexts that shift rapidly and unpredictably make improving and sustaining more difficult. This reality often is advanced as a reason for keeping social program models general. The rationale is that a more generally defined program will not only suit more local situations but will be easier to revise in response to the contingencies that always accompany the passage of time.

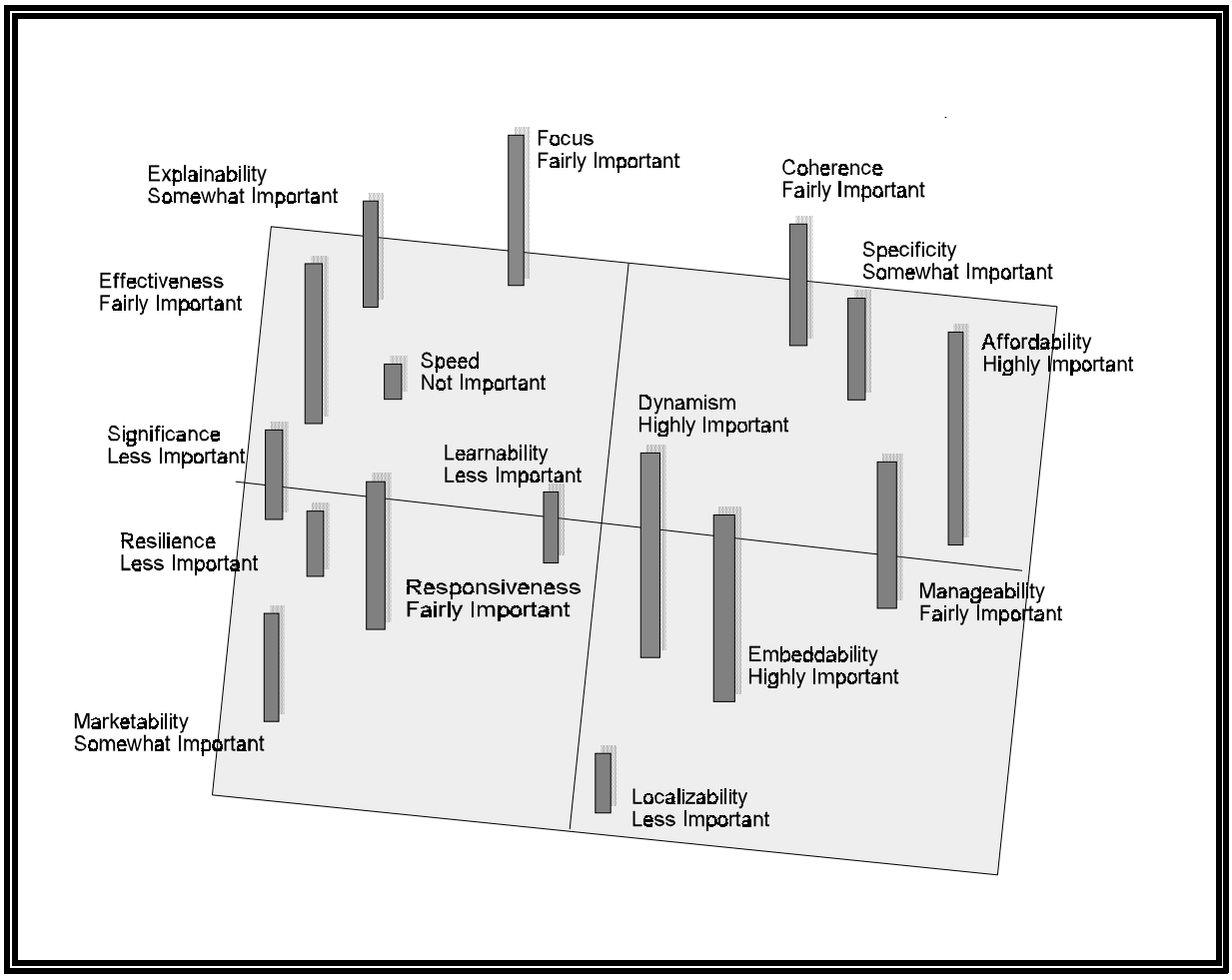
However, just as with localizablity, there can be too much generality or ambiguity in design, such that a program may evolve in quite different ways from one place to the next. So long as the evolution leads to continuing effectiveness for the program overall, there would be no need for concern. But how likely is it to lead in that direction? If the starting point is a general design, there is likely to emerge in the early stages a significant degree of variation in what the program looks like across the settings adopting it. While this variation may produce some superb instances of the model, the probability of these "better" versions strongly influencing the broader network of adopters is apt to be low. Without powerful incentives and the constant circulation within the network of rich, persuasive information on program operations and performance, adopters will be inclined to stick with their own versions. Unfortunately, social markets have shown little ability so far to generate the kind of incentives and information flow that would lead a decentralized development and diffusion process to work as well as it otherwise might under market-like conditions.

A dynamic program is one that is able to maintain an appropriate balance over time between its essential and adaptive elements. It retains specificity where specificity is what

it needs in order to be successful where it exists, and it retains flexibility on the same basis. The elements identified in either category may change with time — indeed, should change. But there is always a balance of some clear sort that enables the program to maintain its integrity and identity in all or most of the settings that have adopted it.

Applying the Dimensions

The dimensions of replicability work in two practical ways. In effect, they call for two kinds of judgment. One kind is establishing, in the case of any given program being considered for replication, the importance of each dimension. A dimension's importance will vary depending on the problem at which a program is aimed and the broader market in which it must operate. For example, although embeddability is always desirable, it will be a more significant dimension for programs that need to function in highly institutionalized markets, such as public education, where fit is both difficult and necessary. It will not matter as much in less institutionalized, more fluid markets — for example, job training for the disadvantaged — where open niches may be more readily available. Or take the dimension of effectiveness. Again, while any program in line for replication should be effective, how effective it must be, and the level of evidence mustered to demonstrate its effectiveness, will be higher in more developed fields like health care than in less developed ones like volunteerism. The importance of each dimension to replicability can be mapped in a three dimensional diagram depicting the"landscape" the program must traverse if it is to be scaled up successfully (see figure following).



The second judgment that the dimensions facilitate involves how a given program rates on each dimension. By matching this second judgment against the assessment of importance, the landscape in the diagram changes to reflect the actual strengths and weaknesses that replication would have to address. The most replicable programs will generally be those that score well on the dimensions expected to be of the most importance in their replication. Conversely, a lack of correspondence between the two judgments will be a strong indicator that a program is a poor candidate for expansion.

Programs vs. Organizations

No program, regardless how well it rates on the dimensions of replicability, is self-diffusing. It needs one or more capable organizations supporting its replication, as well as capable local organizations adopting and implementing it. Indeed, many of the necessary organizational capabilities have been alluded to in the above discussion of dimensions. The obvious question is whether the desired impact on social conditions could be achieved by worrying less about how to identify and advance promising programs and more about developing the capabilities of the social enterprises and public agencies that make up the social sector. After all, businesses appear to succeed over time by virtue of their organizational strengths, rather than by virtue of any particular products or services they offer. Doesn't the same logic apply to social organizations?

While social organizations, too, need healthy capabilities, it is unlikely that just developing stronger organizations will get the job done. There are at least three reasons for this. First is the sector's limited capital. Social enterprises and public agencies, as a rule, do not have anywhere near the kind and level of slack resources that private companies can invest in research and development on new product and service possibilities. Without equally well-financed R&D capabilities, it is difficult to imagine that social organizations can, with more regularity than they do now, generate the innovations that will lead to a healthy rate of social progress.

Second, the competitive pressures that drive businesses to invent and innovate — in effect, to better themselves — are considerably muted in the social sector. In the absence of much stronger competition, it seems unlikely that enhancing organizational capability alone would lead to the sort of social improvement or change associated with the best social programs. And while there may be a role for more competition in social enterprise, the sector's nonprofit legal standing and long roots in an ethic of cooperation make the significant expansion of such competition unlikely.

Third, even if the resources for social investment were plentiful and competition were given much freer rein, it is quite possible that society would balk at the kind of social sector this would be apt to produce. While the sector needs to be innovative, it cannot be as dynamic as the commercial economy and still create value for people and communities. Most social improvement strategies need continuity to be effective; lives cannot be changed or improved through the constant churning of programs and services that would occur if the social realm were a lot more like its profit-making counterpart.

The social sector needs to continue to give special, aggressive attention to the development and promotion of model programs to compensate for the limitations that the sector's peculiarities impose on organizational capability as the path to success. To be sure, the sector needs to be guided by a strategy of organizational strengthening. But it also needs great programs around which to build strong social organizations. The challenge is to capture such programs from wherever they arise — academic research, grassroots efforts, projects of established social enterprises — and give them the support they need to gain broader consideration.

Conclusion

Social markets are governed by a more complex logic than commercial ones. Traditional market mechanisms, like price or consumer response, are of limited use in determining which social interventions should or will spread. Decisions about investing in social programs and practices are reached mainly through political means, which place a premium on making sound judgments about the allocation of capital. These judgments are fraught with uncertainties because of the difficulties associated with reckoning the value of social endeavors. Will Program X, which appeared to work well in Community A, work just as well if it is transferred to Communities B through Z? The answer to this or similar questions is rarely obvious or easily reached.

Getting at the true possibilities necessitates a careful assessment of both a program and the larger context in which it is to operate. The dimensions of replicability represent a systematic method for parsing the opportunity that arises when a model appears ready for broader consideration. They provide a framework that both program developers and funding sources can use to increase the chances of making wise investment choices.