

“The Elusive Craft of Evaluating Advocacy”

Steven Teles and Mark Schmitt

Steven Teles
Department of Political Science
Johns Hopkins University
Steles2@jhu.edu

Mark Schmitt
Roosevelt Institute
schmitt.mark@gmail.com

A version of this paper was published in the *Stanford Social Innovation Review* in May 2011:
http://www.ssireview.org/articles/entry/the_elusive_craft_of_evaluating_advocacy/.

Support provided by The William and Flora Hewlett Foundation

The opinions expressed in this report are those of the authors and do not necessarily reflect the views of The William and Flora Hewlett Foundation.

Nothing big happens without advocacy – concerted efforts to direct public policy (or in some cases, private or corporate decisions) toward socially useful ends. Advocacy includes every form of research, activist journalism, persuasion, coalition-building, and public relations, as well as lobbying and political activity. When advocacy efforts succeed, the results can be transformative – improving lives for millions of people, opening the door to new innovations, or raising the baseline for future efforts by donors. Consider, for example, the investments in the capacity of immigrant-advocacy organizations that led a Republican Congress to reverse draconian limits on benefits to legal permanent residents in 1997, the expansion of charter schools and other education innovations, or the recent health care reform. History shows that good ideas like these did not catch on widely just because they worked. They happened because of creative investments in public persuasion, legislative action, and political activity.

While these points are hardly controversial, organized philanthropy, especially on the center-left, took some time to recognize them. Advocacy requires an approach and a way of thinking about success, failure, progress, and best practices that is very different from the way we approach traditional philanthropic projects such as delivering services or modeling social innovations. It is more subtle and uncertain, less linear, and because it is fundamentally about politics, depends on the outcomes of fights in which good ideas and sound evidence don't always prevail. This difference poses a particular challenge in evaluating advocacy efforts, exacerbating the resistance to advocacy from foundations that naturally resist investing in projects that they can't judge as successes or failures.

The classic formulation of philanthropy's role, originating in the era of "liberal consensus" of the 1950s and 60s, assumed that philanthropy would model and test social innovations, and government would adopt the successes. The approach did claim some victories, but proved inadequate to the political climate of the 1990s and since, an era in which government has been less receptive to ambitious or redistributive policy changes. Meanwhile, on the political right, foundations much smaller than those of the center and the center-left had a disproportionate impact because they efficiently concentrated their efforts on advocacy and ideas, rather than services.¹

For all the successes of advocacy noted above, and many others, there are countless advocacy efforts that have not yielded success – at least not yet. Some may be long shots, others may have no chance of success, while others may quietly be on the verge of a breakthrough. As foundations become increasingly comfortable supporting advocacy, they are also asking themselves how to distinguish worthwhile efforts at social and political change from those that are not a good use of philanthropic dollars. Unfortunately, the craft of advocacy evaluation is far behind the needs of grantmakers, with foundational theoretical issues left largely unasked, much less answered.

Foundations, universities, and government have developed sophisticated tools for evaluating service-delivery programs and smaller-scale tests. These methods range from controlled experiments, to the identification of “best practices” that seem to be transferable from one successful program to another, to a more malleable form of evaluation based on assessing the "theory of change" underlying an initiative. The development and implementation of these tools, often on a large scale, constitutes a growing industry of its own.

¹ David Callahan, "\$1 Billion for Ideas," *The Nation*, April 26, 1999.

These sophisticated tools are, we will show, almost wholly unhelpful in evaluating efforts at advocacy. But there are other kinds of tools, with their own forms of rigor, which are appropriate to the peculiar terrain of advocacy. Advocacy evaluation should be seen as a form of trained judgment, rather than a method. That judgment requires a deep knowledge of and feel for the politics of the issues, strong networks of trust to the key players, an ability to assess organizational quality, and a sense for the right time horizon against which to measure accomplishments. In particular, good advocacy evaluation requires a recognition of the complexity of the terrain of politics, which makes evaluating particular projects (rather than entire fields or organizations) almost impossible. To understand the need for a judgment-laden craft of advocacy evaluation, we should begin by understanding the very different terrain on which advocates and service providers do their work.

(I) Advocacy in the Real World

“Advocacy” is a misnomer. Funders do not, for the most part, give organizations money to simply fly the flag or make the case for a particular policy change. Their goal is to change actual social, policy and political outcomes. And ultimately, if not immediately, advocacy efforts must show progress toward those outcomes. But the relationship between the work done as part of an advocacy effort, and the results or signs of progress are complex and non-linear. Similar resources and strategies sometimes generate very different results. Sometimes political outputs are reasonably proximate and traceable to inputs, but sometimes results are quite indirectly related and take decades to come to fruition. Some advocacy efforts have a specific goal in mind, but in other cases the objective is far broader and benefits are realized by groups other than those who paid the costs. Any effort to evaluate advocacy must be able to account for these and other complicating features

of the terrain of policy and institutional change, but these facts in themselves can't help us evaluate advocacy. A few short case studies will highlight the singularity of each success or failure.

Consider first the campaign for health care reform. The effort that culminated in 2010 was the work of decades, including a previous, high-profile failure in the early 1990s, multiple waves of state-based reform and numerous incremental efforts at the national level. Advocates invested hundreds of millions of dollars on initiatives ranging from media initiatives such as encouraging television producers to include stories of the uninsured, multiple coalition-building projects, university- and think tank-based research and well-funded grassroots initiatives.² The basic outlines of reform policies were worked out well in advance, in advocacy groups and think tanks, who delivered a workable plan to presidential candidates. Key interest groups who could block reform, such as small business, had been part of foundation-supported roundtables seeking common ground for years. Technical problems had been worked out. And tens of millions of dollars had been set aside as long ago as 2007 for politically savvy grassroots advocacy initiatives targeted at key legislators. After a very long slog, the outcome was the Affordable Care Act.

On financial reform, the path was totally different. Few advocacy organizations or think tanks were looking closely at the fragile financial system before it collapsed. The handful of prominent voices for reform, like Harvard Law professor Elizabeth Warren, were operating without specialized organizational support (the main support Warren got was from general interest liberal magazines like *Democracy*, blogs like Talking Points Memo, and the Roosevelt Institute—none of which were designed to deal with financial reform).³ An advocacy organization, Americans for Financial Reform, was created in haste in 2009 and never funded at anything approaching the level given to health care reform. The legislation was by its very nature far too complicated to

² Mark Schmitt, "The Obstacles to Real Health Care Reform," *The American Prospect*, October 26, 2009; Jacob Hacker, "The Road to Somewhere," *Perspectives on Politics*, Fall 2010, pp. 861-876.

³ Mark Schmitt, "Machinery of Progress," *The American Prospect*, December 21, 2009.

mobilize significant public activism, and no tangible constituency (such as, in the case of health reform, the uninsured) would see a specific benefit. Banks, credit card companies and financial firms opposing key elements of reform were well funded and had allies in both parties. As different as the process was, the outcome was similar—the passage of major, compromised but still dramatic, legislation.

Consider by contrast the effort to pass legislation to control global warming, which in many ways closely resembled the strategy to pass health care reform. Advocates of “cap and trade” engaged in what can only be called a mammoth effort, over more than a decade. Among other things, environmentalists drew on the services of a former vice-president who made an Oscar-winning movie, spread their message for more than a decade across a remarkable span of media (up to and including Saturday-morning children’s cartoons), corralled a wide range of well-funded environmental groups to support a single strategy for reducing carbon (cap and trade), and attracted substantial support from large businesses. The movement to control global warming used every trick in the book (as well as creating some new ones), and the result was legislation that never made it to the floor of the Senate, with the very real possibility that action will be delayed by years, if not decades.

A more successful, and totally different, effort concluded in December, 2010, with the repeal of the "Don't Ask, Don't Tell" policy that had barred openly gay men and women from military service. Advocates identified early on a particular elite group whose willingness to vouch for the viability of change would eventually help to shift the opinions of lawmakers. By slowly and carefully cultivating a group of retired officers, and convening roundtables and research to show that gay servicemembers could serve openly without disruption, advocates created the cover that allowed lawmakers to safely change their views on repeal. While grassroots activism, and high-profile maneuvers such as Lt. Gen. Dan Choi's White House protests, helped draw public

attention to the injustice of DADT, the quiet nurturing of elite support was unarguably a key factor in the eventual repeal of the law, just a month after the Senate had blocked the repeal.

The story of efforts to claw back the application of the death penalty played out in a much different way.⁴ From the late 1960s until the late 1990s, net public support for the death penalty (the percentage of the population supporting it minus the percentage that opposed) went up from the single digits to over thirty percent. Executions skyrocketed, from zero in the early 1970s (in the wake of the Supreme Court's *Furman* decision) to over 100 a year by the late 1990s. In response, a small group of church and student groups in the 1980s began to shift their focus from the morality of the death penalty to its actual application, leading to what Frank Baumgartner and his co-authors have called the "discovery of innocence." The effects took some time to come to fruition, but over the last decade they have been dramatic. Net public support for the death penalty dropped in half from its height, the tone of news coverage of the death penalty turned sharply negative as the number of stories went way up, mentions of the concept of innocence in the media skyrocketed, and the number of death sentences was cut in half. Many of the actual effects of this campaign were quite distant from the start of mobilization and vary quite a bit across states, but they are very real and appear to have developed an unstoppable momentum.

A final story, about the creation and nurturing of the Federalist Society, concerns an advocacy project that was not designed, like the previous four initiatives, to produce a single, specific outcome.⁵ The Society began as a network of right of center elite law school groups who seeking to open up space for conservative legal ideas, and over time has developed chapters at every law school in the country, eighty-eight lawyer chapters, and over a dozen practice groups organized around different areas of law. The Society has actively made a point of not involving

⁴ This paragraph is based on Frank Baumgartner, Suzanna De Boef, and Amber Boydston, *The Decline of the Death Penalty and the Discovery of Innocence* (Cambridge University Press, 2008). A shorter version of the argument can be found here: http://www.unc.edu/~fbaum/Innocence/Baumgartner_Innocence_NCOBPS_2008.pdf

⁵ Steven Teles, *The Rise of the Conservative Legal Movement* (Princeton University Press, 2008), Chapter Five.

itself directly in activities where its impact could be easily measured (such as judicial appointment campaigns and litigation) and spends most of its resources on bringing speakers (many of whom disagree with one another) to its chapters. No one doubts the organization's impact, but most of that impact is indirect, through the creation of "movement public goods": the human, social and cultural capital of its members. Put another way, the Society has had a major impact on American law, but most of its benefits have been harvested by other organizations, whose impact has been magnified by the Society's investments.

(II)

Advocacy and Services Are Radically Different - And Their Evaluations Should Be, Too

Foundations that support advocacy have often done so as an outgrowth of their support for service delivery. But the cases discussed in Section I hint at the profound differences between services and advocacy. In advocacy, well-designed efforts often fail, scaled-up efforts often have no more success than smaller ones, and replication of previously successful models doesn't always lead to success. The beginning of wisdom, therefore, is a recognition of how profoundly different advocacy is from services. Evaluation of advocacy must not only recognize these differences, but absorb them in a methodology that is as reflexive and adaptable as successful advocacy efforts are themselves.

Most of the Time, Nothing is Happening

Unlike services, where some progress, however small and atomized, should be visible every day (even if it's just one person escaping hunger for that day), in an advocacy initiative, most of the time very little is or at least seems to be, happening. At any given time, most issues are not even on the agenda of the political system, and energy and money has drifted to those that are. Sociologist

Verta Taylor pointed out in her classic long-term study of the women's movement that unless the fires are kept burning when a cause is in “abeyance” (a process that sometimes lasts decades) and little visible progress is being made, it is much harder to get things moving when conditions become more permissive.⁶ The work of Frank Baumgartner and Brian Jones has shown that these long periods of non-agenda status can be broken quite violently, and often without warning.⁷ If work is not being steadily done during the abeyance period (such as the expert analysis, coalition-building, and legislative design work on health reform undertaken during the 2005-2008 period), then opportunities may be missed or at least left relatively unexploited when the political weather changes.

In services, it is relatively easy to establish benchmarks to measure progress, such as number of persons served, percentage of target population reached, or standards of service quality. But the chaotic, non-linear character of the agenda-setting process means that funders cannot pretend to know where they are at in the process of policy change, so they cannot reasonably evaluate abeyance activity by results or progress toward it. Instead they need to find ways to evaluate abeyance activities on their own terms, not in terms of incremental progress toward a fixed end.

Evaluation of advocacy projects too often assume a linear and steady progression toward success, and advocacy strategists, conditioned by funders, are accustomed to presenting a plan of action in which a large change is preceded by interim goals and achievements along the way. So, for example, a plan to achieve nationwide reform on a key issue might have as interim goals the passage of ballot referenda in several states, a specified number of cosponsors for legislation, or passage of an incremental reform.

⁶ Verta Taylor, “Social Movement Continuity: The Women’s Movement in Abeyance,” *American Sociological Review*, 1989, pp. 761-75.

⁷ Frank Baumgartner and Bryan Jones, *Agendas and Instability in American Politics*, 2nd Ed. (Chicago: University of Chicago Press, 2009).

An organization that can present a plan for advocacy in terms of a well-marked path to success seems like a business with a coherent business plan pointing to profitability, and thus the safest bet for strategic grant-making. And with a plan laid out, a project can be evaluated as to whether it is on track (that is, achieving its interim goals) or off track.

But successful advocates know that such plans are at best loose guides, and the path may branch off in any number of different directions. Interim achievements can be idiosyncratic victories; incremental legislation often satisfies legislators that they have dealt with a problem while exhausting the capacity of grassroots advocates to keep pushing forward; early under-the-radar successes may mobilize the opposition. Indeed, the competitive nature of advocacy, in which opposition continually adapts to block previously successful moves, makes evaluation by timetable particularly dangerous. Worst of all, organizations that are committed to following a predetermined path, especially if they believe their funding depends on it, will be hesitant to adapt to changing circumstances by embracing new opportunities or acknowledging that previously predicted opportunities have been foreclosed.

Successful advocacy efforts are characterized not by their ability to proceed along a predefined track, but by their ability to adapt to changing circumstances. The most effective advocacy and idea-generation organizations, such as the Center on Budget and Policy Priorities, are characterized not by a road map toward a single measurable goal, but by a general organizing principle that can be adapted to many hundreds of situations. Rather than focusing just on an organization's logic model (which can only say what they will do if the most likely scenarios come to pass), funders need to devote most of their time to determining how it will react to unanticipated

challenges or opportunities. That is, the real object of initial evaluation should not be an organization's strategy, but its strategic capacity—its ability to read the shifting environment of politics for subtle signals of change and opportunity, to understand the opposition, and its deftness in devising and executing appropriate adaptations.

Opportunities - and Work - Occur on Many Levels

Our system of government is characterized by parallel, loosely coupled agenda-setting processes at work simultaneously at different levels of government and across institutions. This loose coupling of institutions means that, in sharp contrast to services, advocacy projects cannot realistically experiment in one place in the hopes that successes can be “scaled up.” Successful advocacy projects must simultaneously pursue opportunities at the state, local and federal level, as well as across different governmental institutions. Sometimes these efforts need to be organized into a well-coordinated network across places and institutions, especially when the timing and sequence of action matters. But in other cases they are best left uncoupled to one another, pursued as a portfolio of distinct bets on the assumption that donors have little or no idea which strategy is likely to be successful. Under such conditions, it makes sense to engage in spreading philanthropic dollars across different projects with entirely different theories of change.⁸

Successful efforts to change public policy often require both grassroots as well as elite strategies, often because of the synergies between them, or because opposition in either quarter could derail the idea. (For example, decades of work within the medical profession built elite support for comparative effectiveness standards to ensure appropriate treatment, but with no

⁸The concept of spread-betting is described in Teles, *The Rise of the Conservative Legal Movement*, Conclusion.

grassroots effort, it was easily mischaracterized as "death panels.") In other cases, it is unclear which mechanism or venue of influence will be effective.

Building advocacy projects that cover the full range of opportunities means that, by definition, advocacy work requires massive amounts of effort that will seem wasted, in that it turns out not to have been essential to the final outcome. This waste, however, is fundamentally unavoidable, because neither funders nor the organizations they support can know which strategy will be effective ahead of time. And unlike wasted resources in services, which effective organizations squeeze out over time, an investment in advocacy that seems to have wasted resources may be precisely what turns out to be most important the next time the game is played. In short, in advocacy there is no single "best practice" that can be expected to reliably produce similar results over time (as the comparison between the mobilization efforts in health care and climate change shows).

Spillovers Between Projects Are Pervasive

Especially when viewed in the context of finite funding, there can be a tendency to view issues and advocacy efforts as if they are in competition for a limited amount of political capital or public attention. A focus on child poverty, for example, is treated as an alternative to a focus on retirement security, and success on one would come at the expense of success on the other. This would be true of service delivery – choosing to invest in an early childhood education program is a choice not to invest in a program for seniors. But it is often not of the case in advocacy, where success on one issue often builds success on others—by creating a sense of political momentum, by restoring faith in government, or by creating habits of cooperation within legislative institutions that lead to other successes.

Issue domains that may seem quite distinct in a donor's mind are, therefore, rarely so in the actual world of politics. The pervasiveness of spillovers across issue domains⁹ is one of the reasons why particular issues are almost impossible to disentangle from general ideas that weave issues into broader governing philosophies, and that provide the glue for long-term political coalitions.¹⁰ Consequently, the fortunes of issue-specific mobilization may be due to actions conducted within that domain, but they may be the result of the presence or absence of active mobilization in another domain entirely, or of generic, ideological activity that cuts across issue domains.

It is difficult to accurately attribute the success of any advocacy project to a particular organization (or even issue-specific network), since much of its success or failure may be due to the presence or absence of spillovers from other issue areas, or the presence of generic, cross-cutting activity. External effects of organizational activity are pervasive in advocacy in a way that they are not in services, and thus evaluation is faced with the sticky problem of capturing all the benefits that an organization is generating (as well as preventing it from taking credit for benefits that are in fact produced by others).

The Problems Fight Back

Most service-delivery programs face problems that aren't aware of the efforts to solve them. Hunger, for example, doesn't actively try to derail feeding and nutrition programs, try to cripple the efforts of food banks, or adapt to anti-poverty strategies in order to more effectively keep children hungry. But advocacy efforts almost always involve a fight against a strategic adversary capable of

⁹ The concept of spillovers across policy domains is discussed in John Kingdon, *Agendas, Alternatives and Public Policy* (Longman, 2003), Chapter Eight.

¹⁰ The way that economic ideas influence group interests and coalition-building is discussed in Mark Blyth, *Great Transformations: Economic Ideas and Institutional Change in the Twentieth Century* (Cambridge, 2002) and Cornelia Woll, *Firm Interests: How Governments Shape Business Interests on Global Trade* (Cornell, 2009).

learning and adapting over time. In some cases those counter-strategies come from interests who benefit from things as they are and resist change. In other cases, such as efforts to increase investments in health or education, the main obstacle is competition for scarce resources, rather than outright opposition.

Even areas of public policy that are positive sum in the aggregate are zero-sum where organized interests are concerned—someone loses, and someone wins. The potential losers, consequently, have an incentive to figure out how to make yesterday’s best practice no longer work tomorrow. The declining efficacy of best practices occurs either because the losers figure out a way to adopt the winner’s strategy (there is no IP protection on political strategies), or they find an effective counter-strategy.¹¹ There was a time when bombarding the Hill with phone calls was a very effective way of exercising influence, but it became gradually less so (approaching zero) as everyone else did it. Strategic litigation was a genuinely disruptive innovation in the 1970s, but declined in its impact as its targets developed their own organizations, and as they figured out ways to effectively push back against public interest lawyers. The iterative, competitive nature of advocacy means that it is rarely possible to evaluate advocacy strategies against the metric of best practices.

The competitive nature of advocacy severely limits the utility of logic models as a tool for evaluators. Intelligent military strategists have long taken to heart Moltke the Elder’s observation that, “No campaign plan survives first contact with the enemy.” Any strategy is contingent upon certain expectations of how the adversary will respond. In services, it is usually important that a project be implemented as planned, in order to ensure a reliable test of the treatment, without which no knowledge of broader application can be generated. But the strategy layed out in a logic model for zero-sum forms of advocacy decays from the day it is put into practice, as the adversary

¹¹ The declining efficacy of political strategies is discussed in Teles, *The Rise of the Conservative Legal Movement*, Conclusion.

learns, discovers the weakness of the strategy and devises counter-measures. Logic models may provide an initial sense of whether an organization understands what it is up against, has a coherent strategy and the means to act upon it. What really distinguishes one group from another, however, is what cannot be captured in a logic model—the nimbleness and creativity an organization will display when faced with unexpected moves by its rivals or the decaying effectiveness of its key tools. Whatever utility they may have, logic models in advocacy should not be mistaken for what an organization will actually do in practice.

Systemic Forces and Accidents Matter As Much As The Actions of Organizations

The most important advances in program evaluation have been in the creation of experimental or quasi-experimental designs, which help to ensure that an evaluator is actually measuring the impact of the program under study, rather than the characteristics of the target population or factors outside the program. Attribution is still complicated in practice, but real progress in evaluating services has been made over time.

Attribution in advocacy evaluation, however, runs up against some profound obstacles, the most important being profound structural forces and the density of organizational activity around any issue. As Frank Baumgartner, Jeff Berry and their co-authors have demonstrated, in the overwhelming percentage of cases organizations fail to get substantial traction on their agendas for policy change.¹² The American political system is profoundly wired up for stasis, and competition for limited agenda space is fierce. The failure of a group to get their issue on the agenda or legislation passed should be understood to be the norm even for the most effective and well-designed advocacy organization. Conversely, the random, chaotic quality of what gets on the

¹² Frank Baumgartner, Jeffrey Berry, Marie Hojnacki, David Kimball, and Beth Leech, *Lobbying and Policy Change* (Chicago: University of Chicago Press, 2009).

political agenda means that policy success may have little to nothing to do with the advocacy campaign for it. The fact that health care passed and global warming did not may have been a function mainly of dumb luck and the sequencing of issues on the Congressional agenda, and thus does not reveal anything about which advocacy campaign was better designed or worth funding.

If it is hard to know whether advocacy activity played any part in a particular policy outcome, it is harder still to know whether any particular organization or strategy made the difference. Shifts in agendas often display profound “tipping” structures, where nothing happens until enough energy or activity builds up, leading to an explosion of action. If that is true, the proximity of organizational activity to action tells us nothing about its contribution to the final outcome—the early stages, when nothing was happening, may have played a greater aggregate role in the outcome than the actions of latecomers who claim to have pushed an issue over the finish line. Political success, in short, does not imply organizational effectiveness, and clearer causal connections to the final outcome do not mean a greater contribution to it.

Advocacy is Politics

In other contexts, there's a simpler term for what we're calling advocacy: politics. Every effort to change public policy is political, even if it is an attempt to work around the deficiencies of existing political institutions, such as by forming an independent commission. And thus, evaluating advocacy is complex and nuanced in the same way that evaluating any political effort would be. The points above would not be surprising to a political consultant or journalist trying to explain, for example, why a particular candidate won one election but lost another.

Foundations, however, are somewhat constrained in their ability to evaluate politics because, mostly for appropriate reasons they can't fully engage in politics. They cannot back

candidates or parties, engage in lobbying, or sponsor ballot initiatives. A successful advocacy effort will often, however, involve all these electoral strategies. It is easier, for example, to persuade a legislator to do the right thing if you have some realistic ability to threaten her reelection, such as by putting out a scorecard of votes.

Foundations engage in politics only in ways that are limited by the tax code and by their own sense of appropriateness. For example, foundations often create *bipartisan* commissions or advocacy projects, but are often hesitant to back more clearly partisan initiatives. Internalizing a constraint on partisan behavior, however, can lead to the conviction that bipartisanship is a goal in itself or always a condition for advocacy victory – which it often is, but very often is not, as with health reform. Or, it can lead to a tendency to frame issues in ways that lend themselves to the bipartisan-commission approach, such as on reducing the federal budget deficit, even if that's not the best public policy to achieve the foundation's goals.

When foundations evaluate their advocacy efforts, they often look just at the narrow subset of quasi-political activity that they are able to support, judging the whole effort based on their part—which is not necessarily the most important or effective part. While we do not argue that foundations should press the limits of the law constraining political activity, evaluators still have to watch it intelligently from the sidelines, recognizing the strengths and failures of all the actors in an effort, and thus helping foundations know when their unique strengths and capacities are most helpful. The difficulty that foundations have in evaluating the totality of political activity should also make them hesitant to take a central, coordinating position, a role better left to those with no constraints on seeing the whole terrain of politics.

(IV)
Evaluating Advocacy in an Unpredictable World

The field of advocacy evaluation is full of false analogies and potentially dangerous techniques. But the enterprise of evaluating advocacy is not hopeless. Given the amounts of money that donors are investing in advocacy, it would be foolish and irresponsible to treat it as beyond the scope of evaluation. Evaluation, in some form, is also inevitable—donors need some sort of guiding principles or rules of thumb to decide where to put their money, and when to pull the plug on their existing investments.

Innovation, Not Imitation

It is tempting to evaluate advocacy by drawing on "best practices"—practices that experience shows lead to results. So, for example, a successful advocacy effort that featured an aggressive, populist grassroots campaign might be taken as evidence that such strategies are generally effective. A behind-the-scenes, cross-partisan strategy involving paid lobbyists to contact key swing legislators that achieved legislative victory might be taken as evidence that that the approach works. To the extent that advocates can show they are following previously successful tactical paths, they can claim they are on the road to success.

But as the examples in Section I show, tactics and organizational forms that worked in one instance are not necessarily more likely to succeed in another. Any given advocacy tactic, whether it focuses on the deepest grassroots or the airiest elites, can show successes and failures. A careful

examination will show more failures than successes--something that those who put bread on their table by selling a particular technique will be hesitant to advertise. There is no all-purpose political tactic. What matters, and what funders need to figure out how to evaluate, is the capacity to act strategically, to choose the tactic that is appropriate to the particular terrain of a particular conflict, and to adapt to the shifting moves of the opposition.

In addition, much of what makes a strategy effective is implementation. The fact that one organization succeeded with a particular strategy may be a function of the enthusiasm or skill with which it was implemented, rather than its general applicability. And even if a strategy does have a claim for general applicability, it does not mean that another organization, with very different skills and reputation, can get significant results out of it. The “best practice” in politics is matching the right strategy to the problem at hand—something that can usually only be known with certainty after the fact, if at all.

In other cases, advocacy projects are judged not by their similarity to other successful strategies, but to organizational structures that are seen as successful. Liberals used to say, with tiresome frequency, "We need a Heritage Foundation for our side," but since the Center for American Progress came to full fruition, now conservatives say, "We need a CAP for our side." Unarguably successful organizations such as the Center on Budget and Policy Priorities, are often imitated in form, when the more useful lesson from their success is the adaptability of their form.

By focusing on replicating existing models, advocacy organizations, and their evaluators, often fail to step back and ask the question of what kind of organization is most appropriate to the particular challenge. In one example of organizational imitation, The American Constitution Society was driven by the perception that the left needed a response to the Federalist Society. Consequently, funders helped build an organization that mirrored the Society's structure, with

student groups, lawyer's chapters and practice groups, with a headquarters in DC. Mimicking the Federalist Society sidestepped the question of whether the proper response was creating an organization that resembled legal liberals' rivals, or an entirely different structure designed from scratch. ACS might turn out to be successful at pursuing the organizational tasks it mimicked from the Federalist Society, but that does not mean that those are the tasks that are actually the most vital ones to pursue.

Because opponents of change are structured and prepared to contest existing advocacy models, a more effective effort might come from a disruptive innovation – a strategy or organizational form that does not follow known strategies. Consider, for example, MoveOn.org, which has become one of the most effective multi-issue advocacy organizations in the U.S., but which for many years after its establishment in 1998 was doubted because its model of repeated, cheap small actions by members was so different from the model of one-time organizational membership that was considered the test of effectiveness. Just as important, it may take time for disruptive innovators to find the most effective application of their tactic, just as new technologies only generate productivity gains after a long period of trial and error. Consequently, an evaluation of an advocacy innovation might show it as having relatively little efficacy early on, but large returns in later periods—but a head-to-head comparison of a new approach would show it to be less effective than an older road-tested strategy.

Further reinforcing the case for valuing disruptive innovations – and thus unfamiliar models—the policy process is not a static environment. What worked yesterday may not work tomorrow. Even as recently as twenty-five years ago, when the Tax Reform Act of 1986 passed, the policy process at the national level was characterized by extremely weak parties, strong committees and sub-committees in Congress, significant room for bureaucratic and interest group entrepreneurship, and pervasive cross-party coalition-building for major reforms. A set of best

practices for reform based on that policy process, focused on building expert consensus and cross-party “strange bedfellow” coalitions would have little relevance for today’s national policy process, characterized as it is by polarized, highly disciplined parties. Best practices are a function of time, and they may also be a function of issue area—what works in one domain may be very different from what works in another. Education, for example, is still characterized by significant cross-party coalitions, while global warming, as we noted in Section I, has gradually been pulled into the maw of party competition.

Finally, political strategies have a sell-by date. In direct services, it is not usually the case that there is an active, creative opposition trying to make your strategy fail. But as we argued in the previous section, that is exactly the situation in large areas of public policy. Tactics for public policy change depend, in many cases, on catching an adversary unaware, using an approach that it finds difficult to understand, or exploiting an organizational weakness that makes it hard for them to respond effectively. Almost all of these advantages wear out over time, as one’s adversary learns and adapts. In fact, adopting a “best practice” can even become a disadvantage, if it means that one’s moves are easily predicted and countered. Politics, like war, rarely stays at equilibrium, and so success requires constant innovation to keep one’s adversary off-balance and force it onto the defensive.

Assess the Entire Advocacy Infrastructure, Not Single Projects

Despite the number of organizations that will present themselves as the decisive force behind any legislative accomplishment, no successful advocacy effort is the result of any one organization or initiative. While Health Care for America Now was a large, visible, and indispensable force behind passage of health care, when the full story is told it will include

countless other efforts, such as the Herndon Alliance of 2005-2006 and many others. Some of those efforts were far from government, such as the academic work at Dartmouth that showed how health care cost growth could be contained while improving services, and some that weren't directly focused on health care at all, such as political organizing around a broad progressive agenda and candidates. Much of this work was not funded by, or could not be funded by, foundations, but was nonetheless vital.

The last point is particularly important: A successful advocacy infrastructure is rarely limited to safely fundable 501(c)3 projects. To take the clearest example, initiatives that depend on ballot initiatives involve (c)3 work that doesn't count as lobbying for research and generic advocacy, followed by c(4) activity once signature-gathering begins. A heavy investment in the first stage that cannot be matched by an equally aggressive formal campaign will fail, and the first investment will have been wasted. Even though foundations do not want to get involved with campaigns they cannot fund, they need to know about them and evaluate them alongside the things they can fund. This is only the starkest example of a situation in which the entire organizational infrastructure needs to be assessed, not just a single grantee. Even organizations that one does not like or trust are part of the overall advocacy infrastructure that has to be understood.

To evaluate an entire field, donors must ask whether there is an adequate balance of grassroots and elite work, whether soft policy work can be followed up with hard political work, whether organizations are duplicating work and creating confusion, and whether state-level affiliates or allies are adequately supported or just a thin network of allies in name only. The “advocacy field” also includes organizations doing work broader than the particular issue that the donor is particularly focused on. For example, a donor concerned with Middle East peace or poverty reduction needs to recognize that their issue is embedded in a much broader context. The attitudes of the public and elected officials on these issues are as influenced by efforts to legitimate

broader, ideologically-inflected concepts (such as the proper role of the United States in the world and beliefs about the American social contract) as they are by issue-specific mobilization.

Evaluating the success of the latter requires taking into account the efficacy of the former, whether donors are directly funding them or not.

View Advocacy Grantmaking as a Portfolio of Bets

Looking at an entire field calls for a "spread betting" approach to advocacy, and thus to evaluation. A foundation investor should care not about the success or payoff of any one grant, but the aggregate payoff of their entire portfolio of investments, relative to their costs. Coupled with a respect for disruptive innovation, portfolio investment implies pushing evaluation to the broadest possible scope of activities. An investment in an issue in which no action has occurred, even for a long time, may not be a bad use of resources. But this will only be clear when a particular issue is judged in the context of a range of other bets put down by the donor. Only then can a donor have a sense of whether their resources are generating what investors call "alpha"—excess returns over the average. Portfolio evaluation, by averaging out a number of investments over a longer period of time, also prevents the risk of over-attribution of success or failure to factors that are entirely exogenous to the activities of those they are investing in.

A spread betting approach recognizes the importance of the structural features of the political system—its random, chaotic character—and that success or failure is as much a matter of the luck of the draw as the skill of particular actors. Given that, it makes sense to invest for a very wide range of different scenarios (including some that may have a low probability of paying off, if the potential returns are large enough), rather than focusing their investment just on those that seem most likely *ex ante*. At the broadest possible level, spread betting means investing in a range

of different issues (subject to the donor's ability to effectively understand and supervise them), and not just different strategies within a particular issue.

Spread betting is an approach to managing risk appropriate to the world of uncertainty we have described throughout this paper. It recognizes that failing to fund the seemingly quirky, unproven strategy that turns out to be appropriate to the circumstances is just as big a risk as funding something that turns out not to work out. Spread betting also draws attention to the risk involved in putting all of a funder's resources down on a single organization or strategy. While that single organization or strategy may be less risky than the alternatives, it may be that the combined risk of a collection of more unorthodox approaches may actually be lower in the aggregate.

Spread betting, therefore, requires that funders have an internal organizational culture that can accept a considerable number of failures, so long as they are balanced over the long term by a few notable successes. What matters, in short, is the performance of the philanthropic portfolio as a whole, not the success or failure of any one particular investment.

Adopt a Long Time Horizon

While spread betting and portfolio evaluation point to the importance of assessing investment across a range of issues and strategies, our next principle suggests evaluating investments using the longest feasible time horizon. First and foremost, the longer the time horizon a donor applies to his portfolio of investments, the greater her ability to control for the random qualities of the policy process and identify activities that are actually generating value that would not have existed in their absence (rather than just being in the right place at the right time, or vice versa). In addition, since funder time horizons necessarily influence the activities of the groups they

fund, a longer time horizon will avoid creating an artificial incentive for organizations to pursue short-term (and usually lower return) strategies over those with a longer payoff period.

Second, and just as important, a long time horizon recognizes that the political process does not end after a piece of legislation passes, a regulation is promulgated, or a court decision is handed down. As Eric Patashnik has argued, such an approach would ignore the critical importance of “policy durability”—whether a reform actually sticks, or even creates a platform for further change.¹³ Some reforms, such as airline deregulation and tradeable permits for sulphur dioxide, generated powerful reinforcing dynamics that kept the policies from being clawed back, even in the face of initially strong opposition. But other changes that seemed momentous at the time, such as the Tax Reform Act of 1986, came unraveled bit by bit over the years, until much of their original force was diminished. Viewing policy enactment as only one step in a much longer process focuses donors’ attention on the fact that what really matters is whether a policy sinks deeply into society and political routines, not whether a piece of legislation passes. Funders may not be able to wait for years after reforms have been passed to judge whether their investment in producing them was worth it. But at the very least they should consider the possibility of reversal (or extension) in their evaluations, and evaluate the strategies of advocates by whether they have a plausible plan for protecting what they have won in the ballot box, legislature, courtroom or regulatory agency.

Pay Attention to Policies’ Impact on Politics, Not Just on Target Populations

A longer time horizon can allow funders to detect positive as well as negative feedback from policy change. Some policy changes matter because they change the playing field on which subsequent action can occur. For example, the state welfare reforms of the 1980s and early 1990s

¹³ Eric Patashnik, *Reforms at Risk* (Princeton: Princeton University Press, 2008).

(many of which were supported by donors like the Bradley Foundation) did much more than change policy in the places where experiments were carried out. The reforms altered the entire debate on the issue, and reduced the uncertainty that faced policymakers when they tried more ambitious changes at the national level.¹⁴ In some cases policy feedback is more diffuse. On the negative feedback side, Suzanne Mettler has shown that the shift of federal support for higher education to tax subsidies (as compared to the direct, visible mechanism of Pell Grants) led citizens to fail to recognize the degree to which they are recipients of federal support.¹⁵ By contrast, the direct, traceable structure of the GI Bill had powerful positive effects on citizen's attitudes toward government. Andrea Campbell found similar positive mobilization effects from Social Security, whose structure "fed back" into senior citizens' high levels of political participation and programmatic knowledge.¹⁶

In many cases, policy changes have political feedback as one of their primary objectives. Investment in "green jobs," for example, was expected to have an impact far beyond the actual creation of jobs or improvement to the environment: it would create a lasting labor-environmentalist alliance, mobilize voters around an optimistic economic vision, put a bright face on the tough choices of cap-and-trade or carbon taxation, and might even create a national security message of reduced dependence on oil from the Middle East. Similarly, much investment in children's issues saw them as a strategic wedge to build a more positive view of government, as children were seen as more legitimately in need of public help. Libertarian litigation on issues like school choice and property rights was pursued in part on its own merits, but also because of its potential to reduce public (in particular racial minority) perceptions of conservative and

¹⁴ Steven Teles, *Whose Welfare? AFDC and Elite Politics* (University Press of Kansas, 1998).

¹⁵ Suzanne Mettler, "Reconstituting the Submerged State: The Challenges of Social Policy Reform in the Obama Era," *Perspectives on Politics*, Fall 2010, pp. 803-824.

¹⁶ Andrea Campbell, "Self-Interest, Social Security, and the Distinctive Participation Patterns of Senior Citizens," *American Political Science Review*, September 2002, pp. 565-574.

libertarians' good will on racial issues.¹⁷ In the long term, those on the right hoped that this sort of activism would “detoxify” their brand, opening the way for both electoral and policy inroads that would be blocked so long as their side was viewed as being tarred with the brush of racial animus.

These negative and positive feedback mechanisms can be very powerful, but they often play out very slowly. In many cases, the long-term effects of policy change (which often operate through their influence on other policy domains) on the character of politics are at least as important as those that are produced by the policies themselves. While many of these effects are far after a grant has been closed, donors who fail to at least consider their importance (and encourage their grantees to take them into account) will be ignoring some of the most significant, transformative effects of public policy. And in doing so, they may shorten the time horizons of their grantees, creating powerful incentives to make decisions that favor strategies that generate short-term policy gains at the risk of long-term policy reversal and minimal alterations to the long-term effects on political attitudes and mobilization.

Evaluate Organizations as a Whole

Many of our points so far have been about the proper “unit of analysis” for evaluation. We have suggested a bias toward a unit of analysis that is long-term, portfolio-based, and inclusive of diffuse and indirect effects as well as those that are direct and easily traceable. This points to perhaps our most radical suggestion, which is that funders may be better off eschewing evaluating *advocacy* (in the sense of discrete activities) at all and instead focus on evaluating *advocates*. That is, the proper unit of analysis is the long-term adaptability, strategic capacity, and ultimately influence of organizations themselves. This is the grant-making model of, among others, the Sandler family's work helping to create the Center for American Progress and Pro Publica, the

¹⁷ Steven Teles, “The Eternal Return of Compassionate Conservatism,” *National Affairs*, Fall 2009, pp. 107-126.

Olin Foundation from the 1980s-2000s, and the Walton Foundation in its work on education.

While the fundamental epistemological issues associated with evaluating particular instances of advocacy approach the impossible, many of those problems are less severe when the unit of analysis is shifted to the organization as a whole.

Many advocacy organizations can themselves be conceptualized as spread bettors. That is, they engage in multiple projects at any one time. The best organizations can be conceptualized as the ones that, over the long term, generate outsized returns as compared to their competitors. In the short term, the success or failure of an organizational portfolio is likely to be due to the vagaries of forces outside its control (such as swings of the electoral pendulum). But the longer the time horizon, the more these short-term forces wash out and the organization's ability to generate what investors call excess returns (or alpha) reveals itself.

Evaluating organizations also means attending closely to the value they generated for others, rather than only focusing on its direct impacts. In Section One, we pointed to the example of the Federalist Society and its creation of "movement public goods."¹⁸ This would be the extreme case of an organization that exists primarily to generate increase the productivity of others. But even organizations whose main product is something else may generate quite substantial external benefits, or spillovers. For example, certain magazines (one can think of *the Washington Monthly*, the *American Prospect* or, in another era, *The Public Interest* as examples) have acquired a reputation for identifying good young writers (often those who had been overlooked before), training them, giving them opportunities to acquire a reputation by writing ambitious pieces for a broad audience, and associating them with the magazine's prestigious brand. These magazines regularly lose their staff to other publications further up the journalistic food chain. But what is a loss for the magazine is a gain for the larger ecosystem of which it is a part. Evaluating organizations

¹⁸ Teles, *The Rise of the Conservative Legal Movement*, Chapter Five.

therefore means assessing the value that they add to others, not just what is generated by their own work product. The job of the advocacy evaluator is knowing enough to recognize whether staff turnover is a desirable product of organizational success, or—when leaders are abusive or just simply not good at getting value out of their staff—of organizational failure.

In many cases where influence is extremely diffuse, such as think tanks, magazines and some university projects, it may be pointless to try to trace any particular organization input to a policy outcome. The best an evaluator can do is try to judge the actual content of the organization's activities directly, by (in the case of think tanks, magazines, and scholarly research) reading and making a judgment call on its quality, or by in other ways directly observing and assessing the merits of its core organizational activities. This was the practice of many conservative foundations, whose staff devoted much of their time to being intelligent consumers of their grantees' work products, trusting that their judgment of good, appropriate work would ultimately be vindicated in the real world of politics.

If donors cannot easily determine which strategy will pay off on the front end, or measure outputs at the back end, they should focus on what they can know at least something about, which is the strategic capacity of an organization. Strategic capacity is in part a function of an organization's senior leadership, but as Marshall Ganz has argued, it is also a product of how an organization thinks and acts collectively.¹⁹ A good organization has a coherent and inspiring internal culture, the ability to consistently identify and motivate talented people, deliberate as a group, acquire intelligence on the environment and process it intelligently, and to devise sometimes unexpected responses and to effectively coordinate its actions. In addition, good organizations have the ability to innovate and reorganize when their key tactics have grown stale or have been proven

¹⁹ Marshall Ganz, *Why David Sometimes Wins: Leadership, Organization and Strategy in the California Farm Worker Movement* (Oxford, 2009).

to be ineffective—in short, to learn and reallocate capital internally in response to feedback. What donors should care about is strategy as an evolving feature of an organization, and these indicators of strategic capacity are the right place to start.

The final way to measure organizational quality and influence is through “network evaluation”—figuring out what other organizations and influential individuals in its policy space think of it. While this is probably the most important form of knowledge, it is also the most difficult for a foundation to discover in a reliable fashion. In cases where organizations are in competition with each other for resources (even, or especially, when they are on the same side), peer evaluations may be far too harsh. When organizational leaders have close personal links, by contrast, their assessments are likely to be too generous. And of course all advocates have profound incentives to overstate their own importance in a policy community (since creating the perception of influence may actually produce influence), and to draw funders’ attention to selective information designed to make their organization seem indispensable.

In most cases, participants in a policy network may be hesitant to share accurate information with outsiders with whom they lack ongoing relationships, such as survey researchers or consultants hired by the foundation. Especially when the information is about their own organization, advocates will only reveal challenges or obstacles to those with whom they have profound levels of trust and durable relationships. Despite these problems, members of policy networks generally do develop reasonably accurate assessments of the quality of their peers—of who they listen to and trust, who does good work, and who policymakers take seriously. The challenge is to create an evaluation system capable of penetrating the inherently non-transparent character of political action, and extracting the “private knowledge” of political actors.

When trying to assess influence, what donors are really looking for is “network centrality”—what actors play vital roles in issue networks, in some cases operating as connectors between

communities that lack other relationships or are characterized by mistrust. The network maps produced by, for example, Chris Ansell and Sarah Reckhow in their recent study of Bay Area education groups, are one systematic attempt to demonstrate networks of influence, including funders.²⁰ But the kind of information used by Ansell and Reckhow can only get at certain superficial (albeit important) forms of connection, not actual influence. The real challenge for funders is figuring out how to get this more vital, but hard to find information. How they might do that is our final subject, to which we now turn.

(V)

Conclusion:

What Does a Good Advocacy Evaluator Look Like?

If the argument above is correct, then the core challenge in advocacy evaluation is access to quality information. Most of what a funder needs to know about an advocacy organization or initiative exists in the minds of people in the evaluated organization's network, but the incentives for them to pass it along to an evaluator in pure, unadulterated form are weak, and in some cases negative.

In advocacy evaluation, the "cleaner" the data the less likely it is to measure what matters. Good advocacy evaluation typically involves using trained judgment to acquire and accurately weigh and synthesize otherwise unavailable but imperfect information, from biased and self-interested sources with incomplete knowledge under rapidly changing circumstances where causal links are almost impossible to establish. These conditions mean that advocacy evaluation should not aspire to being formalized into a method, in the sense of a series of techniques that could be replicated by any individual with the same training. Advocacy evaluation is, instead, a craft, one in

²⁰ Andrew Kelly, Chris Ansell and Sarah Reckhow, "How to Reform a Reform Coalition: Outreach, Agenda Expansion and Brokerage in Urban School Reform," *Policy Studies Journal*, 2009, pp. 717-743.

which tacit knowledge, skill and networks are more useful than the application of a rigid methodology. It is an exercise in trained judgment.²¹ The evaluator, rather than the formal qualities of the evaluation, is what matters.

If scientific method is inappropriate, where can grantmakers look for an analogy that can shed light on the intensely judgmental quality of advocacy? One suggestive model is the skilled foreign intelligence analyst. She consumes official government reports and statistics—which she is aware provides an only partial picture of the world, and one with very significant gaps (often areas that would make reveal the malign intent, corruption or incompetence of the government). She talks to insiders, some of whom she has learned she can trust, and others whose information and opinions she has learned must be taken with (at least) a grain of salt. In many cases, she learns as much from what she recognizes to be lies as from the truth. She is aware that those she is trying to learn about are aware that she is trying to understand what they are doing, and thus have an incentive to spread misinformation and engender mistrust in all her sources. But she is also independent of the diplomats and other agents who are trying to effect change themselves.

It is the web of all of these imperfect sources of information—instead of a single measure—that helps the analyst figure out what is actually happening. And it is the quality and experience of the analyst—her tacit knowledge—that allows her to sort through all this information and create a picture of what is happening.

The best intelligence analysts are really applied anthropologists. They study a particular culture, in a particular place, when how it actually works is different than what is on paper, or even how people describe it. Cultures are often characterized by a “hidden structure” that is largely invisible to outsiders, and sometimes poorly understood even by insiders. Many cultures actually

²¹ An excellent discussion of the concept of judgment, and the dangers when it is squeezed out of the process of evaluation (in this case of finance) can be found in Amar Bhidé, *A Call for Judgment: Sensible Finance for a Dynamic Economy* (New York: Oxford University Press, 2010).

develop a considerable lack of transparency, precisely to prevent comprehension by outsiders, who would use that information to tax, regulate or police them. Discovering how a culture works requires “participant-observation” and encouraging networks of “informants.” This requires the generation of trust, which may take years to develop.

What one pays an intelligence analyst for, and what one should pay a good grantmaker in the field of advocacy, is those relationships, the ability to penetrate intentionally or unintentionally opaque surfaces, and the ability to detect patterns of influence that are usually difficult to discern from the outside. Those networks and tacit knowledge are the core assets of a foundation engaged in advocacy, which suggests the need to build this capacity internally, to strive for substantial continuity (and thus institutional memory) among those who possess it, and to defer to their trained, subjective judgment in making key decisions.

The characteristic features of the terrain of politics—chaotic agenda-setting, pervasive deception and misinformation, overlapping, shared responsibility—means that no one metric can capture the reality of influence. Donors do themselves a disservice by even looking for one. It is only by trying to make sense of policymaking activity through the simultaneous application of multiple ways of knowing that donors can get closer to finding out what they need to make their consequential, judgment-laden, but unavoidable decisions.