

Afterword

A BIPARTISAN MONEYBALL AGENDA⁶⁰

By **Robert Gordon**

Former executive associate director at the Office of
Management and Budget

Ron Haskins

Senior fellow at the Brookings Institution

*We asked two scholars, **Robert Gordon**, a Democrat who worked in the Obama White House, and **Ron Haskins**, a Republican who worked in the Bush White House, to put together a broad agenda of policy proposals for Moneyball. And we asked them to keep a key question in mind as they developed it: What are the most practical and the most bipartisan steps we can take at the federal level to make Moneyball our reality?*

Over the last two decades, government has made real progress in playing Moneyball. Some programs now tell potential grantees that they need to bring the evidence if they're going to get the money. A few agencies support research institutes that fund, analyze, and publicly present rigorous research. Social programs now use far more data to support their efforts than even a decade ago.

But use of evidence and data still isn't the norm for government. When we go into the drugstore and pick up some medicine, we know the company that manufactured it has conducted gold-standard research to demonstrate that the drug is safe and, at least in important cases, effective. When we log on to Amazon, we know that they are using enormous amounts of data to tailor their suggestions to our interests. (How we feel about that is a different issue.) But when we interact with a typical government program, we don't expect—and don't receive—the same effective and refined use of research.

History shows that government can raise its game. It used to be that a new mayor or governor or president could hire or fire staff at will based on party loyalty. The creation of the civil service, whatever its flaws, ended that practice. It used to be that government officials who expensed personal items would usually get away with it. Improvements in accounting, together with independent watchdogs like inspectors general, have curtailed those abuses. Pharmaceutical companies always test drugs because the law requires it.

New laws and organizations can help increase the use of evidence, but they won't be enough. There must be changes in leadership and culture. In this afterword, we outline some ideas for this transformation. They fall into three categories:

- Building leadership and infrastructure
- Transforming federal programs to learn and fund what works
- Moving funds from less-effective to more-effective programs

These three nuts-and-bolts proposals cover a lot of ground, but they are united by four pillars that should guide all decisions at every level of government. Here they are:

Pillar 1: Relentlessly use data and evaluation to learn from experience. Without a way of identifying what works and what doesn't, progress in social policy is impossible. Until recently, the most sophisticated evaluations required a lot of time and money. Sometimes that's still true, but not always. With modern data systems, we can do quick, sophisticated tests of different program designs. Think about a store chain testing different product placements in different stores—or a social-services agency testing

different intake routines in different offices. To figure out cheaply what works, we can often use data that governments already collect. Think about a new math textbook: rather than setting up a whole new approach to collecting data, we can just assign the book to half the classes (selected at random) in a district and compare the scores of kids who used the new text with the scores of those who didn't, on tests the kids already take. And once we learn the best interventions, we can subject them to financial analysis to compare benefits and costs—and thus give policy makers an important tool to help make tough choices about different ways to spend limited resources.

Pillar 2: Define success in terms of measurable, transparent outcomes. Programs need goals for improving the world—not just spending money, providing services, or following rules. To summon public pressure for results, there should be clarity and openness about what these goals are and whether programs are achieving them.

Pillar 3: Create incentives to do what works. In a well-functioning market, companies that satisfy their customers and turn a profit grow, while less successful firms either improve or die. Sometimes public-sector programs can successfully simulate this environment by conditioning public funding on outcomes. This is often the right way to go. Yet there can also be practical challenges. Success may only become clear over a long period (think of preschool programs that aim, among other things, to increase college enrollment and completion), while funding decisions

have to be made now. If too much depends on outcome metrics that lack great sophistication, gaming the system becomes a problem. For these reasons, rather than tying too much to measurable outcomes, it often makes sense to incentivize evidence-based program designs that are likely to achieve those outcomes.

Pillar 4: Foster innovation by cutting mandates and using waivers. Good people trying to get good results in diverse circumstances will often figure out better solutions than Washington can. The federal government needs to overcome its habit of viewing long rulebooks as paths to program improvement. Once there are good outcome incentives in place, hard-and-fast rules should be linked to essentials for protecting health and safety, avoiding the misappropriation of funds, or guaranteeing a national floor of benefits or services. Sometimes eliminating requirements for everyone and putting faith in local flexibility and measurement of outcomes will make sense. Other times, where accountability is weaker, it may be wiser to waive program rules for selected states or grantees that agree to adopt stronger metrics in a particular context.

RECOMMENDATIONS

Now to our recommendations. Just a note about scope: We focus on social interventions that are central to expanding individual opportunity. But we believe that these recommendations have application to most domestic policies—certainly to efforts aimed at creating jobs, for example—and to some foreign policies as well, notably foreign aid.

BUILDING LEADERSHIP AND INFRASTRUCTURE

Recommendation 1: Create a chief evaluation officer of the United States—and for every agency. Just as the Oakland Athletics wouldn't have succeeded at Moneyball without Billy Beane, government's efforts won't succeed without strong senior managers who have the power to make and back up decisions. There are too many pressures on decision makers that go beyond focusing on outcomes. So the executive branch needs a leader and small staff to captain Moneyball initiatives, and then each agency needs the same kind of team.

This does not require a bigger bureaucracy. The Office of Management and Budget (OMB) already works to drive evaluation work, but the office has an immense range of mandates, and the Moneyball agenda rises and falls with different directors. Talented career staff contribute a lot, but many of them have other “real jobs.” To institutionalize Moneyball, OMB needs a chief evaluation officer who can provide expert advice to the OMB director and senior agency leadership on how to integrate research into decision mak-

ing and who can coordinate data, evaluation, and performance management to strengthen agency capacity for learning. This would combine, in one elevated role, two separate jobs in recent years: the executive associate director and the associate director for performance management.

Like OMB, most agencies today have evaluation units, but some lack sufficiently senior leadership. In others, out of an admirable desire to ensure that research isn't politicized, key evaluators play only a modest role in designing programs—including their evaluations. But policy makers and program designers need input from evaluators to do their jobs well. And each agency needs a chief evaluation officer reporting to the secretary or deputy secretary.

Agencies can make changes to get this done. In recent years, the Department of Labor did not have a strong culture of evaluation. But in President Obama's first term, the department's leadership hired distinguished chief evaluation officers, first Jean Grossman from Princeton and then Demetra Nightingale from the Urban Institute. Deputy Secretary Seth Harris required every bureau to work with the chief evaluation officers and develop a learning agenda to identify what works in priority areas. Today evidence and evaluation are parts of every discretionary grant program at the department. Moreover, the department initiated two major evidence-based programs and one of the first federal Pay for Success initiatives. If other federal agencies had similarly empowered leaders, they could produce similar results.

Recommendation 2: Set aside up to 1 percent for evaluation at each agency. According to a recent GAO report, only 37 percent of government program managers said that an evaluation of their programs had been completed in the last five years. That's disturbing but not surprising. Historically, evaluation has often been lumped with program administration that nobody likes to fund.

This attitude is understandable but misguided. The private sector spends well over 12 percent of all its domestic investment on R&D, and the amount of investment has been increasing since 1950. But at agencies like the Department of Education, R&D is well under 1 percent of discretionary spending.

To make sure every agency can access the funding it needs for evaluation, with the flexibility to spend it well, agencies should have authority to direct up to 1 percent of their total discretionary funds toward program evaluation. While many agencies would choose to spend less in some years, the authority would help ensure that the agency did not miss important learning opportunities when they arise. Agencies could draw the funds from across each discretionary program on an equal basis and then spend on the highest-priority evaluations, subject to congressional oversight. The Department of Labor now has an authority along these lines, and other agencies should get it, too.

Recommendation 3: Create cross-government prizes for innovative approaches to evaluation. Agencies also need incentives to commit more energy to evaluation. Early in the Obama administration, OMB ran a competition for the best evaluation ideas, then put winning proposals into its budget. The competition stimulated excellent new thinking but lost momentum after Congress failed to appropriate funding. More recently, the Coalition for Evidence-Based Policy has demonstrated the power of a fully funded prize to elicit new ideas for evaluations: \$300,000 in prize money this year drew fifty-three prize proposals for low-cost randomized controlled trials (RCTs) in important policy areas. While the coalition has adequate funding to award only three prizes, some of the runners-up are likely to be funded by philanthropy or government.

Building on these examples, Congress should appropriate a small amount of funds for OMB to allocate for critical advances in evaluation. In some instances, OMB could help agencies strengthen their capacity for low-cost randomized controlled trials and rapid-cycle evaluations, especially if administrative data can be used as outcome measures. In other cases, OMB could guarantee funding for evaluations that answer critical questions that drive policy decisions. An OMB-run allocation process would provide added support for evaluators within agencies, ensure that cutting-edge initiatives receive funding, and provide a forum for shared learning across and beyond government. Indeed, the review of applications could be a public conversation with experts outside government that eventually attracts philanthropic and university support.

Recommendation 4: Create comprehensive, easy-to-use “what works” databases at each agency—and develop a gateway for searching them all. It’s not enough to evaluate programs if nobody knows about the results. The evidence about what works needs to be available in a single place, online, in a format that is true to the research but accessible to laypeople, especially policy makers and their staffs. Every government-funded study needs to be accounted for, including studies that are ongoing, so there is no question whether results are being suppressed. And such an Internet source should set high standards before suggesting conclusions.

Putting evaluations online, with high evidence standards, can inform better decision making and signal to researchers the importance of using rigorous research and evaluation designs. Before the What Works Clearinghouse (WWC) of the Institute of Education Sciences (IES), there was a powerful tendency in education to blur the distinction among types of research. An RCT using validated performance measures might be viewed as no more meaningful than an after-the-fact survey of program participants. These would be “two conflicting studies.” The WWC lifted standards and clarified which programs were supported by rigorous research evidence and which were not. There is still real work to do—particularly in emphasizing the importance of effects. But the WWC is real progress.

In the last few years, other agencies and programs have followed suit, creating similarly comprehensive online resources in areas as diverse as home visiting and job training. But there are still no “what works” databases for huge swaths of federal programs,

including much of health, housing, and national-service initiatives. Even among agencies that have established clearinghouses, none identifies all government-funded, ongoing, or completed research online in one place. And the databases are separate, even though researchers and practitioners often work across issue areas.

Most important, information is not always accessible to practitioners and actionable on the ground. There need to be more step-by-step guides showing people how to apply successes from elsewhere in their circumstances, or, even more effective, access to coaching, webinars, or other tools to help community and nonprofit leaders understand successful implementation. This will require a new level of collaboration among policy makers, program managers, and evaluators. Government should set aside a little funding, then partner with philanthropy to facilitate the development of cross-subject platforms, user-friendly websites, and high-quality support for implementation. The Pew-MacArthur Results First Initiative is a great first step here and should serve as a model for other foundations.

Recommendation 5: Create and institutionalize “Moneyball scores”—showing not just what works, but what works best for the money. Using evidence is a solid step forward, but policy makers and practitioners also need to know if the program they are evaluating works better for the money than the alternatives. This is called cost-benefit analysis or “return on investment” analysis.

How do we quantify the benefits from improving a child’s reading, or stopping a violent crime, or saving a life? These are fraught

questions. But we're answering many of them today in other contexts. Since the Reagan administration, a part of the Office of Management and Budget called the Office of Information and Regulatory Affairs (OIRA) has estimated the social costs and benefits from different regulations. While policies can advance human dignity in important ways that cost-benefit analysis doesn't capture, OIRA is often able to apply a simple principle: a regulation shouldn't go forward unless the benefits exceed the costs. Over the last ten years, officials in both Democratic and Republican administrations have reported that regulatory efforts have produced hundreds of billions of dollars in net benefits.⁶¹

While there are many challenges, a similar process could apply to federal grants. The Robin Hood Foundation, which fights poverty in New York City, already uses cost-benefit analysis to compare interventions across diverse domains. And the Washington State Institute for Public Policy has developed cost-benefit estimates for a range of interventions, from reducing criminal recidivism to expanding access to early-childhood education. The Washington State legislature—and now, legislatures in about a dozen other states, thanks to the Pew-MacArthur Results First Initiative—is using cost-benefit analysis to consider whether its own social investments are getting results.⁶²

Because of the complexity and controversy inherent in cost-benefit analysis, the federal government will need a nonpolitical expert entity to design principles with substantial public input. The Institute of Medicine, with funding from the MacArthur Foundation,

is now convening an expert panel to develop common principles for analyzing prevention programs. While there will never be universal agreement on methodology, either that body or another should continue and broaden its work until there is a widely endorsed approach to producing estimates (or ranges of estimates) of costs and benefits for a broad array of social-intervention programs.

At that point, both executive and congressional analysts should begin to employ cost-benefit approaches in providing Moneyball scores for legislation. Nobody does that today. The Congressional Budget Office focuses on the costs of legislation but only sometimes looks at benefits. For example, CBO will show savings to the government in a program that is permanently funded but not one that is funded each year. CBO's mandate should be broadened—or a new office created—to report on the evidence supporting different initiatives and to consider the broad costs and benefits from government expenditures.

Executive agencies should also create Moneyball scores to evaluate different grantees. In addition to determining whether a program is using an approach that gets results, an agency would determine how the cost per outcome in that program stacks up against other programs serving similar populations.

Recommendation 6: Build crosscutting data systems that also protect privacy. The key to Moneyball for government is data—especially low-cost data. In recent years, governments at all levels have been increasing their investments in building systems to collect data, protect it, and match it across domains (education,

labor, health, etc.). The Obama administration has created a default policy of making government data sets available to the public. Many cities are building their own open data platforms.

But there is still a great deal of work to do. Too often, key data systems are not connected. For example, evaluations of postsecondary-education training programs would ideally reflect baseline information about high school achievement, as well as subsequent employment outcomes. Evaluations of elementary-school interventions would similarly draw on background data from social-services systems. Yet this is often impossible because of policies restricting data sharing.

One problem here is cost. Early in the Obama administration, Congress invested strongly in strengthening longitudinal data systems in both education and worker training. But as discretionary spending has come under tight caps, this funding has largely dried up.

While it makes sense to press harder for data-infrastructure funding, such an effort will run into real concerns about data privacy. Such fears sank *inBloom*, an effort to create an easy way for teachers to see the critical information about their students. The lesson here is that policy makers cannot expect to build data systems, show they work, and hope everyone will thank them. Instead, they need to address data issues up front, as systems are being built, in a way that conveys that the gains in efficiency from using data won't crowd out people's concerns about their dignity and privacy.

Representative Paul Ryan (R-WI) has a proposal for a Commission on Evidence-Based Policy Making with the charge to decide whether and how to bring together through a single clearinghouse multiple data sources—and to settle on the privacy protections for that database. Putting privacy issues at the center of the conversation can help ensure that the smart use of data gains broad support.

Recommendation 7: Build human capital in government. While *Moneyball* uses big data and advanced statistics, the endeavor ultimately depends on people. To take one example: rapid-cycle evaluation at a service-delivery agency requires policy experts to identify plausible new approaches; program managers to implement them; technologists to create or modernize data systems to capture effects; social scientists or statisticians to analyze the data on effects; and crosscutting leaders who know how to bring these pieces together with inspiration and precision.

Within government, increasing use of short-term fellowships and assignments from academic institutions can help build temporary expertise. But government needs outstanding individuals who dedicate many years to the work.

Unfortunately, governments at all levels increasingly struggle to get and keep the people they need. In part due to repeated pay freezes and shutdown crises, federal-employee job satisfaction is at its lowest level since 2003.⁶³ Attrition has risen 37 percent since 2009. While public-sector jobs always pay less than their private-sector counterparts, the gap is widening.⁶⁴ For employees with advanced degrees, average pay is now 18 percent below the private

sector.⁶⁵ Attracting and retaining excellent staff is made more difficult by rules governing hiring, retention, and promotion that were developed for an industrial economy seventy years ago.

The Partnership for Public Service and others have proposed comprehensive reforms in the civil service, including changes in pay, hiring, and retention practices. These recommendations—and comparable efforts at the state and local levels—deserve a close look.⁶⁶

TRANSFORMING FEDERAL PROGRAMS TO LEARN AND FUND WHAT WORKS

Recommendation 8: Protect, improve, and grow tiered-evidence programs. Building on a Bush administration concept, the Obama administration created six tiered-evidence programs across the federal government.⁶⁷ These programs represent a breakthrough because of two design features: First, they provide more money to programs with higher levels of evidence (hence the “tiers”), creating a strong incentive to do what works. Second, they require evaluations so that programs can continue learning and improving.

Even though these types of initiatives represent some of the best in government, their survival is far from assured. Neither of the evidence initiatives at the Department of Labor still exists in its original form. The key spending committee in the House of Representatives has previously targeted all of the programs for elimination. And because these programs are all creations of the Obama administration, they run the risk of being forgotten by the next president in either party regardless of their merits.

If we want a government that works, stepping back from tiered-evidence initiatives would be a terrible mistake. A recent review of the evidence-based education initiative noted that three of its four largest grantees have established strong positive outcomes in rigorous evaluations (and the fourth evaluation has not yet been completed). Across the six original initiatives, there are now around seven hundred programs being carried out, most with evaluations and many with RCTs. These will soon produce a whole new generation of evidence about what works. Other evidence-based initiatives for parental home visiting, teen-pregnancy prevention, and the Social Innovation Fund (SIF) have funded programs that have promising records of success.

As a more general matter, there's good reason to believe that at current funding levels, the evidence-based programs have a higher return on investment (ROI) than typical government programs. For example, based on nonexperimental research, the average federal job-training program has a modest positive ROI.⁶⁸ But job-training programs vary widely. Some are terrific, some are terrible, and most are in the middle. By prioritizing approaches proven to work, evidence-based workforce programs are likely to achieve higher average levels of impact. And the requirement to evaluate results will provide a basis to improve programs all along the spectrum of effectiveness.

To be sure, evidence-based initiatives still have room to improve. One recurring error is treating interventions as proven when their effects are *statistically* significant, even when some of

those effects aren't so important for taxpayers. For example, take three home-visiting programs: One of the programs reduces infant mortality, a second reduces child abuse, and a third increases the number of referrals made to other providers. The first two programs get much more important results than the third, but currently the law treats all three programs as equally strong. That should change.

Even with improvements, the evidence initiatives have room to grow. In 2013, the Investing in Innovation Fund at the Department of Education had 618 applicants and made twenty-five grants, giving it an acceptance rate comparable to a highly selective college.⁶⁹ Many unfunded applicants received scores that were trivially different from winners. Similarly, evidence-based home-visitation programs currently serve a small fraction of low-income mothers who are eligible. Right now, it makes sense not only to protect these evidence-based initiatives but also to expand them.

Recommendation 9: Grow Pay for Success and scale what works for social mobility. While tiered-evidence initiatives require past records of success, they provide taxpayer money even if programs fail. A different approach, Pay for Success, goes a step further. Rather than pay for a service up front, the government enters an agreement to pay only after a program delivers specified results, usually results that save taxpayers money. Providers raise money from philanthropic or private-sector funders and participate in a third-party evaluation. If the program achieves its targets and generates savings, the government returns a portion of those sav-

ings to the funders. In effect, the promise of future savings funds preventive services right now.

This approach to funding resembles investing in a new business. If the business turns a profit, the investments pay off; if not, all or part of the investment is lost—to the investor or philanthropist, but, in this case, not the taxpayer. An especially desirable feature of Pay for Success is that it brings many of the benefits of business investing to government programs: clear goals, important outcome measures to judge success, rigorous evaluations of program outcomes, and pressure on program operators to produce impacts.

To date, several agencies have supported small Pay for Success programs. But there's the potential to do more, especially in the areas of educating young children with disabilities and reducing homelessness. Recently, bipartisan teams in the House and Senate have introduced bills to provide a permanent source of funding for Pay for Success, a \$300 million fund based at the Department of the Treasury, which can fund interventions across different agencies.⁷⁰ Congress should enact these proposals.

As exciting as Pay for Success is, it is still a new and complex idea, and it can be a poor fit for Washington in cases where federal taxpayer savings do not exist or arise over the very long term. Intervention targeted at young children without special needs, for example, can have long-term effects on education and earnings but probably won't yield rapid financial savings. These long-term outcomes need to be encouraged, especially because they can contribute to restoring America's promise of intergenerational social

mobility. The tiered-evidence-based initiatives sometimes create that opportunity, but it's spotty. For example, the Department of Education's Investing in Innovation program began with the idea of "scaling up" what works, but in recent competitions, it has aimed to spread more limited funds across more providers with less evidence. The Social Innovation Fund has similarly focused on funding promising ideas.

For these reasons, we recommend widening the Pay for Success initiative to offer substantial up-front grant funding for the handful of programs that are demonstrated to have a high social return on investment and a large impact on increasing social mobility. Few programs would meet this standard at first, but that is okay. Taxpayers—and the people designing and running good programs—ought to know that money is available to grow extraordinary interventions with extraordinary returns.

Grantees would first need to cross the threshold of having rigorous evidence of a large impact on an important outcome, whether it's getting ready for school or finding a good job or avoiding prison. To permit comparisons across issue areas, from preventing recidivism to encouraging college completion, grantees would be ranked using cost-benefit analysis, and only initiatives with high return on investment would be funded. Funding could be provided in a Pay for Success format, but it would not need to be. Program administration could be consolidated with the new Pay for Success fund proposed at the Department of the Treasury, which would be closely coordinated with poverty-reduction

programs at other agencies. Use of intermediaries could be encouraged, but not required as in the Social Innovation Fund, and support could be given to programs that aim to achieve multiple outcomes across a population.

The combination of Pay for Success and scaling “what works” initiatives could generate billions of dollars in new funding for effective interventions. But because those costs are so substantial, we offer proposals for specific funding mechanisms later.

Recommendation 10: Transform existing formula and competitive grants. Our previous three recommendations would expand the most-innovative approaches to funding what works. But these recommendations don’t address an even tougher question: What will we do with the great majority of federal grant programs that *don’t* require evidence? Unless we improve the way these programs allocate funds, the federal government will continue to waste taxpayer money, and many interventions won’t produce the results they could.

Changing core programs—such as the Community Development Block Grant—will take hard work and courage. Programs have built constituencies around one way of doing business. Change will often be unwelcome. But it’s necessary.

Different strategies for basing programs on evidence will be appropriate in different cases. In existing competitive grant programs, there should be priorities for grant proposals that bring strong evidence of important successes and that require evaluations. Because running a grant program that evaluates evidence and

supports rigorous site-level evaluations is costly, increasing use of evidence will sometimes mean consolidating smaller programs. Such consolidation can not only allow for evidence to influence funding decisions but also eliminate inefficient federal rules that get in the way of the best approaches to problem solving. (See our “lookback” discussion below.)

Promoting the use of evidence in programs that make grants based on a formula (usually related to population) is more complicated. The simplest approach is to require grantees to use a portion of formula funds for interventions proven to be evidence based. But this only works where there are strong, accessible lists of practices that grantees have the ability to use. Congress recently used this approach for mental health block grants.

Another promising strategy is to award a small portion of large formula grants to be awarded competitively based upon grantees’ effective use of formula grants—as measured by outcomes, the use of evidence-based practices, or some combination of these metrics. The Bush administration’s \$10 million home-visiting initiative worked this way. The federal government can also require states and localities to use some formula funds to develop their own capacity for evidence-based decision making. And lastly, when metrics are strong, the federal government can simply require that formula grantees meet baseline performance standards or lose funding.

The ultimate goal is that formula grant programs provide so much support for evidence-based initiatives and their evaluation that many tiered initiatives become unnecessary. Little by little, the

evidence-based approach can improve the nation's domestic programs and even lead to continuous improvement through rigorous evaluation. As part of the "lookback" process described below, the next administration should conduct an assessment of major funding streams to determine which strategies to pursue in each context.

Recommendation 11: Conduct a grant-program "look-back" to replace mandates for processes with incentives for outcomes. For many grant programs, simply demanding greater use of evidence and evaluation is not enough. These programs also require all sorts of procedures and services that constrain innovation. At the same time, the programs aren't clear about what they need to accomplish. That structure needs to be reversed.

Consider Head Start, the nation's largest and most important preschool program. In the only large-scale RCT ever performed on Head Start, children who participated in the program saw immediate gains in school readiness, but any advantage appeared to vanish by third grade.⁷¹ Yet other early-childhood programs have stronger outcomes.

Head Start has "performance standards" that run to 145 pages and dictate everything from seating arrangements at lunch to membership on parent councils to timing for parent-teacher conferences.⁷² While any one rule is reasonable, the cumulative effect of having so many is to discourage innovation and encourage a culture of box checking instead. Head Start still lacks a clear, manageable list of the skills (cognitive, social, emotional) that the

program is supposed to impart and measure. Partly as a result, most programs use curricula that have been shown in research not to drive improved child outcomes.⁷³

The people who run Head Start programs are deeply dedicated to children and want to do right by them. And Head Start is improving. It now requires grantees that receive poor ratings to compete to keep their grants. Programs are rated based on expert observations of classrooms, using procedures that have been shown to modestly predict student outcomes. The combination of ratings and “recompetition” represents real progress. But there is a lot of work to do. Even with the competition, most incumbents are keeping their grants—including those who have achieved only mediocre results.

Head Start should develop and implement a set of performance measures that reflect the skills young children need to succeed in school. At the same time, the program should free providers from many of the current performance standards so that they can try different approaches to helping kids learn. The recompetition process should include a broader range of performance measures. And all programs should have new incentives to adopt more evidence-based practices, such as better curricula.

We are confident that these types of reforms would be helpful in a range of federal programs, from Job Corps to college-access initiatives. And the government could use a common process for identifying such reforms. In the regulatory context, OIRA in 2011 led a government-wide regulatory “lookback” process requir-

ing agencies to “reexamine their significant rules and streamline, reduce, improve, or eliminate them on the basis of that examination.”⁷⁴ The lookback process yielded scores of measures to update regulatory regimes.

Virtually every grant program could use such a lookback as well—a review of all the current rules to see if they have the right balance between mandating a baseline of services supported by evidence and incentivizing excellence. In some cases, legislation would be needed, but in others, regulations could achieve significant reforms. The goal of the lookback process would be to institute stronger performance metrics, incentives, and evaluations; to simplify or eliminate accreted requirements that are no longer useful; and, in some cases, to eliminate small, siloed, and prescriptive programs and replace them with larger, more integrated, and more evidence-based initiatives.

Recommendation 12: Create new flexibility to test new approaches to fighting poverty. Evaluation of state welfare experiments led to the important reforms of the nation’s major cash welfare program enacted in 1996. These experiments were made possible by a provision in the Social Security Act allowing the secretary of health and human services to grant waivers that would permit demonstrations designed to promote the purposes of the act. Other major federal social programs should have—and use—similar authorities that encourage experimentation by state and local governments. This would clear the path for new approaches to increasing self-sufficiency and cut poverty.

The problem is that what strikes you as sensible experimentation may strike me as an attack on all that is good in the world. And nobody wants experimentation to be unlimited. But if we are to improve the quality of government, both political parties will need to show greater willingness to err on the side of innovation.

In 2012, President Obama proposed to allow states to change the way they measure the number of people who are engaged in “work-related activities” under welfare. Republicans immediately jumped on this proposal as an attempt to “gut welfare reform.” In fact, the proposal would have allowed states to test plausible new ideas for supporting self-sufficiency, such as greater use of subsidized employment. Republicans had previously supported greater flexibility in welfare programs, and for good reason since the program had been created in 1996 based on experiences with waivers in forty states. Still, the administration’s announcement just before the election was problematic, and the criticism was intense. In response, the Obama administration made some sensible changes to the proposal, but by this point the issue was so politically toxic that nobody was interested in waivers. That’s too bad, and the idea deserves a second look. Waivers from static program rules almost always deserve a second look.

Recently Congressman Paul Ryan proposed what he has called Opportunity Grants. These would allow a handful of states to combine a wide range of both discretionary and entitlement programs, including food stamps, welfare, child care, and housing assistance. Aid recipients would be required to draw up a “life plan” with

social-service providers, and able-bodied people would be required to work. States would contract with different organizations that provide services to the poor. The states and the federal government would set benchmarks for outcomes and evaluate the results.

Many progressives have responded to this proposal by focusing on its most controversial component: the elimination of the individual guarantee of the Supplemental Nutrition Assistance Program (SNAP), formerly known as food stamps. And it is fair to ask how this cost-neutral proposal jibes with House Republicans' other proposals to make significant cuts in SNAP and other social programs. Since welfare reform, SNAP represents the country's only commitment to guaranteeing benefits to all poor individuals and families.

Still, as with the Republican response to Obama's waivers, Democrats should not overlook ways in which the Ryan proposal matches ideas of their own. The government's compartmentalization of poverty programs (housing, education, health, etc.) often doesn't match up with real life. A victim of domestic violence may need emergency housing, transportation to school for a child, and mental-health services all at once. She needs an integrated solution, not a jigsaw puzzle with the pieces thrown all over the floor—different programs, eligibility requirements, offices, and so forth.

In the particular context of helping “disconnected youth” (those aged 16–24 and not in school or working), President Obama proposed and Congress recently created an authority for “performance partnerships” that will allow states and localities to pool annually appropriated funds and receive waivers from program-specific

requirements in education, training, health, and community-service programs. State or local leaders would propose new metrics for improving outcomes, and results would be evaluated.

Congress should come together behind extending the existing performance-partnership authority into new domains, including programs addressing housing and criminal justice. And while it makes sense for progressives to want to preserve SNAP as an entitlement, the other programs that Ryan proposed to subject to waivers do not currently reach all eligible individuals anyway. With strong protections (including for current recipients of these benefits), Congress should offer more flexibility to a few states over these funds, provided that in return they commit to getting better outcomes, as measured through rigorous evaluation.

Offering more flexibility in exchange for more accountability and evaluation will often be a promising formula, though in some instances it will require new funding in order to work. The recently enacted Farm Bill included \$200 million to test out new approaches to providing job training for food-stamp recipients. If these approaches succeed in moving significant numbers of individuals out of poverty, they could save far more than \$200 million in a program that cost more than \$80 billion in 2013. Because of the evaluations, we'll know.

The same approach could also help strengthen some of America's largest and most important programs. For example, the disability-insurance program within Social Security provides a safety net for millions of Americans who have worked hard during

their lives but are no longer able to work due to illness or injury. The program costs more than \$140 billion per year, and the share of working-age Americans who are receiving payments has risen substantially over the last three decades, a trend largely but not entirely explained by demographic and economic trends.⁷⁵ As Jeff Liebman and Jack Smalligan have shown, there are innovations in the program that might help protect the vulnerable, hold overall costs to current or lower levels, and enable more individuals to stay in the workforce.⁷⁶ And the Social Security Administration could be testing these approaches today if it had appropriate authority and funding. Congress should provide them.

Finally, communities could utilize the combination of flexibility and funding to tackle America's most central challenges. Liebman has fleshed out the compelling idea of a "Ten-Year Challenge" that would modestly fund ten communities to seek to achieve breakthrough results in addressing ten problems over ten years, ranging from recidivism to homelessness, with flexibility as needed across funding streams and an evaluation of the results.⁷⁷ State and local governments could achieve a great deal more if they were able to organize their efforts around meeting these ambitious goals, rather than complying with different programmatic requirements.

MOVING FUNDS FROM LESS-EFFECTIVE TO MORE-EFFECTIVE PROGRAMS

Recommendation 13: Offset costs to encourage reforms.

Democrats and Republicans disagree about whether to increase domestic spending. So do the two of us. But we both agree that the

Moneyball agenda we have outlined should be implemented even if domestic spending does not increase by one penny.

From a cost perspective, our proposals fall into three categories. First, some recommendations improve existing programs with no new costs. Reforms to existing competitive and formula grant programs and targeted interventions fall into this category.

Second, expanding the federal evaluation infrastructure would have modest costs. If savings to pay for this spending are needed, it is appropriate to create a small levy across entire agencies. The needed reduction in current spending would be very modest. The cross-agency approach avoids creating a line item that budget cutters could foolishly target. It also reflects the reality that evaluation and data systems provide broad benefits for agencies, their programs, and the populations they are supposed to serve.

Third, a few of our recommendations have substantial new costs. Expanding existing evidence-based initiatives and creating a “Scale What Works for Social Mobility” program could cost billions, though their social benefits are likely to be much greater. When waivers are granted, testing new approaches (as with disability) can require new up-front funding as well.

There are several options for cuts to pay for this new spending. President Obama’s budgets have included several billion dollars in as-yet un-enacted program cuts that could be sources of new resources.⁷⁸ These measures range from reductions in spending on oil-and-gas research to reforms and reductions in the Senior Community Service Employment Program.

If these measures are not sufficient, a final good option would be to reduce the largest block-grant programs that currently lack strong evidence criteria. These reductions could be linked to increases in evidence-based initiatives. For example, if the Scale What Works initiative funded a large evidence-based initiative to scale up a proven job-training strategy, the reductions could come from the job-training formula programs. This is not a first choice, because the formula grants ensure at least some level of services across the country, and evidence-based programs will not replace all services in all places that are lost due to cuts. However, evidence-based programs are carefully designed to achieve higher levels of impact than typical block grants. In addition, because of their emphasis on evaluation, these programs can generate learning that is useful across all government programs. Finally, while reductions in block grants will lead to reductions in services in some places, the increases in evidence-based programs will lead to expansions in services in other places. The greater concentration of services in some areas over others is not optimal, but roughly the same number of people would be served as from formula-only programs. The key is that with the evidence-based programs, the average quality of service and their average impacts on social problems will be higher.

CONCLUSION

If you're a member of Congress or an executive-branch official who is looking to allocate scarce dollars, what do you do? Historically, you make changes on the margins. You look at the research and do your best to infer what works, realizing you don't know much. You try to avoid big political problems.

If you're a nonprofit-organization leader looking for government funding, what do you do? Historically, you find someone who knows someone who makes decisions. You get some constituents to lobby on your behalf. You do some polling. These days, maybe you start a social-media campaign.

These are exactly the dynamics that our proposals would change. They create a vibrant marketplace with both supply and demand in what works. They create the supply of what works by funding more evaluations to identify effective interventions. And they create the demand for what works by driving grant dollars toward interventions that work.

In this new era, government officials will have access to rich data and a far better idea of what works. And program leaders will know that in order to get funding, what they will need isn't influence or anecdotes; it's evidence.

And that's what it means to play Moneyball for government.