

NBER WORKING PAPER SERIES
SCOPE CHALLENGES TO SOCIAL IMPACT

Monica P. Bhatt
Jonathan Guryan
Jens Ludwig
Anuj K. Shah

Working Paper 28406
<http://www.nber.org/papers/w28406>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
January 2021

This paper was made possible by the generous support of the Laura and John Arnold, Edna McConnell Clark, Crown, Lloyd A. Fry, Joyce, Reva and David Logan, John D. and Catherine T. MacArthur, Pritzker-Pucker and Spencer foundations, the Chicago Center for Youth Violence Prevention, EquiTrust, JPAL-North America, the Illinois Criminal Justice Information Authority, the city of Chicago, grant number 2012-JU-FX-0019 from the Office of Juvenile Justice and Delinquency Prevention, Office of Justice Programs, US Department of Justice, and award number 1P01HD076816 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development of the National Institutes of Health. We are also grateful for operating grants to the University of Chicago Crime Lab from Susan and Tom Dunn, Ira Handler, and the MacArthur, McCormick and Pritzker foundations. We would like to highlight the contributions and thought partnership of the Chicago Youth Violence Project Team: Philip J. Cook, Kenneth Dodge, George Farkas, Susan E. Mayer, Harold A. Pollack, Julia Quinn, Maitreyi Sistla, Laurence Steinberg, Nathan Weil, and John Wolf. For vital assistance in making this work possible, we thank Chad Adams, Roseanna Ander, Barbara Byrd-Bennett, Valerie Chang, Molly Clark, Akeshia Craven, Rukiya Curvey-Johnson, Gretchen Cusick, Aarti Dhupelia, Tony DiVittorio, Mayor Rahm Emanuel, Wendy Fine, Jaureese Gaines, Michael Goldstein, Craig Howard, Tim Jackson, Barbara Kelley, Ed Klunk, Timothy Knowles, Tim Lavery, Stig Leschly, Jonathan Lewin, Arnaldo Rivera, Janey Rountree, Alan Safran, Leonetta Sanders, Mai Sistla, Sara Stoelinga, Elizabeth Swanson, Robert Tracy, and Karen Van Ausdal, as well as the staffs of the Chicago Public School system and Youth Guidance. Thanks to Jeffrey Broom, Sarah Dickson, Kylie Klein and Stacy Norris for their help in accessing the data we analyze here, and to Nathan Hess, Stephanie Kirmer, Daniel Kowalski, Khoa Nguyen, Kelsey Reid, Michael Rosenbaum, Catherine Schwarz, Noah Sebek, Robert Webber, Nathan Weil and David Welgus for their invaluable contributions to the data analysis. Thanks to Peter Bergman, Ben Castleman, Josh Goodman, Katy Milkman, Sendhil Mullainathan, Todd Rogers and seminar participants at the Harvard Business School for helpful comments and suggestions. Points of view or opinions in this document are those of the authors and do not necessarily represent the views of the US Department of Justice, National Institutes of Health or any other funder. Any errors are of course our own. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2021 by Monica P. Bhatt, Jonathan Guryan, Jens Ludwig, and Anuj K. Shah. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Scope Challenges to Social Impact

Monica P. Bhatt, Jonathan Guryan, Jens Ludwig, and Anuj K. Shah

NBER Working Paper No. 28406

January 2021

JEL No. A1,Z18

ABSTRACT

Over the past 50 years, social science has increasingly become involved in the business of not just understanding social problems, but of helping solve them as well. Progress towards this goal will benefit not just from the growing work on solving the challenge of scale, by which we mean the capacity to effectively reach large numbers of people, but also from more attention to the less-appreciated challenge of scope, by which we mean the capacity to change a large share of the decisions that affect a given person's outcomes. In this essay, we provide a simple framework to help think about the scope of candidate policies, and highlight some common situations that are often associated with potential for large scope.

Monica P. Bhatt
University of Chicago
Crime and Education Labs
33 North LaSalle Street
Suite 1600
Chicago, IL 60602
mbhatt@uchicago.edu

Jens Ludwig
Harris School of Public Policy
University of Chicago
1307 East 60th Street
Chicago, IL 60637
and NBER
jludwig@uchicago.edu

Jonathan Guryan
Northwestern University
Institute for Policy Research
2040 Sheridan Road
Evanston, IL 60208
and NBER
j-guryan@northwestern.edu

Anuj K. Shah
Booth School of Business
University of Chicago
5807 South Woodlawn Avenue
Chicago, IL 60637
anuj.shah@chicagobooth.edu

A data appendix is available at <http://www.nber.org/data-appendix/w28406>

I. INTRODUCTION

Fifty years ago the Social Science Research Council (SSRC) published an essay by its then-president, the distinguished psychologist Henry Riecken, who argued:¹

“The social sciences, like the physical or biological sciences, are intellectual subjects, directed primarily toward understanding rather than action. It would of course be a curious kind of ‘understanding’ that had no implications for action, and this is perhaps especially true for the social sciences. Nevertheless, there is a difference between enlarging one’s understanding of human behavior and society on the one hand and trying to solve a social problem on the other.”

Since then, the social sciences have increasingly tried to help “solve social problems.” Economics, for example, has become an increasingly empirical field over time (Hamermesh, 2013 and Angrist et al., 2017). In 1974, Nobel laureate Robert Solow and others created the Manpower Demonstration Research Corporation to help implement interventions out in the world as part of a new experimental program to measure their impacts. This was part of the leading edge of the ‘credibility revolution’ that has made empirical work increasingly relevant for policy (Angrist and Pischke, 2010). In 2003, the Poverty Action Lab was founded at MIT with the mission of “working to reduce poverty.” Later, the ‘behavioral revolution’ gave economists new tools to help inform policy design. In 2010, the UK government launched a behavioral insights team – the “nudge unit” – to inform government policy design, followed a few years later by similar teams in the United States and Australia. Economists are now actively working on the challenge of how to scale policies and programs found to be successful in demonstration projects (de Rezende and Eluf, 2016, Al-Ubaydli et al., 2017, Banerjee et al., 2017. Davis et al., 2017, and Muralidharan, et al., 2017).²

We argue here that these efforts would benefit from an addition to the conceptual toolkit we use to judge the potential impact of candidate policies. To see what’s missing, consider how we would decide whether it is worth developing and testing one behavioral science-informed policy intervention: swapping the location of unhealthy foods with fruit and other healthy snacks in school cafeterias. What are the questions we would normally ask ourselves about whether to pursue studying this intervention using our standard economics framework?

- *Is there a real problem here?* Yes. Childhood obesity has tripled in the US since the 1970s, with around one in three children and teens now overweight or obese. There are important externalities and internalities associated with obesity.
- *Does theory suggest a promising policy lever?* Yes. A large body of behavioral research tells us that defaults matter by capitalizing on the fact that switching is effortful, by signaling an endorsement by whoever is structuring the choice environment, and through reference dependence (Dinner et al., 2011).
- *Does the intervention seem likely to pass a benefit-cost test?* Yes. The cost of swapping the location of healthy and unhealthy foods in school cafeterias is negligible. Given the social costs of obesity, even a modest impact would be enough to justify these costs.

¹ Reickan also served as the first director of the National Science Foundation’s social sciences division.

² Public health has long been focused not just on identifying health problems but solving them, starting with the famous identification of a cholera-spreading water pump in London by John Snow in 1849.

- *Will the intervention scale?* Yes. Like many nudges, this intervention is largely mechanical and so easily replicated. We could imagine doing this for every cafeteria serving each of the 32 million children who eat school lunch in America.

This candidate policy checks every one of the boxes on our list. Nonetheless, this policy change would not have more than a modest impact on obesity prevalence.

Where is the breakdown? The following calculation illustrates the problem. Hanks et al. (2012) report that this intervention increases the share of food consumed by children in the school cafeteria that is healthy from 33% to 36%, or three percentage points (about 10% of baseline). Under the usual benefit-cost calculus about whether this is worth doing, we can see that this looks amazing given that the cost of the intervention is close to zero. It would be crazy *not* to do this.

But in a world in which we want to help achieve big changes in major social problems, there is another set of questions we should ask ourselves. Suppose that children eat, say, 25% of their weekly meals at school. Changing cafeteria layout policies would increase children's weekly healthy-food consumption by 3% times 25%, or less than one percentage point. Perhaps the potential impact could be larger if some children eat school breakfast, not just lunch, or if we could enlist cafeterias in other settings like malls or museums. But *eventually* we will run out of easy opportunities to change healthy eating by changing check-out line options – limited by the modest share of eating children do in cafeterias.

The problem with the usual framework is that it ignores the issue of *scope*, by which we mean: of the different decisions an individual makes that affect the outcome of interest, what share of them are affected by a given intervention? That is, how large of an effect could the intervention have for a given person's outcome of key policy concern? If *scale* is the question of an intervention's capacity to reach *across* people and serve an increasingly large population, *scope* could be thought of as an intervention's potential reach *within* a given person across the different decisions that shape the outcome of concern.

In the next section we discuss in a bit more detail what we mean by scope and what features of settings and/or policies are commonly associated with the potential for large scope. The third section discusses some challenges that can arise in trying to solve the scope challenge in practice.

II. ASSESSING SCOPE

In what follows we provide a bit more description of what exactly we mean by scope. Forecasting the scope of a candidate intervention is relevant for the goal of focusing on things that have the potential to have a large impact at the population-level on some policy problem. This is separate from the question of whether a candidate intervention generates benefits to society in excess of the intervention's costs.³ Sometimes it is possible to achieve large

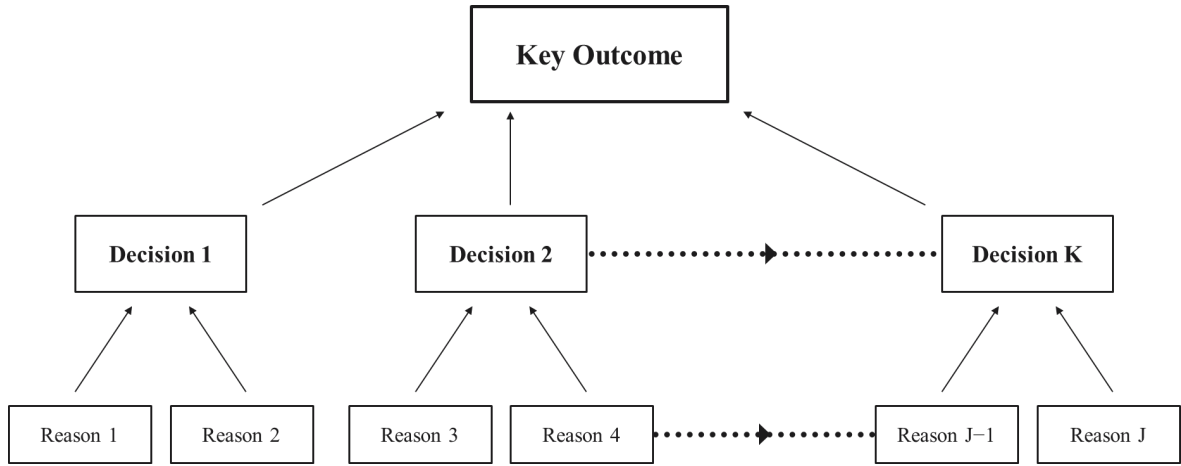
³ Just as a single outcome might sometimes be a function of multiple decisions, it can be that a single decision might affect multiple outcomes. In that case we might redefine what we mean by "large impact on a social problem" to be not just a large change in a single aggregate outcome, but a large change in an index of outcomes where we might, for instance, dollar-weight the index by the value to society of a unit change in the outcome.

population impacts through the aggregation of small impacts from multiple interventions that all pass a benefit-cost test. But because this is not always possible, for reasons we discuss below, it is valuable to assess a candidate intervention’s scope above and beyond some benefit-cost measure.

A. A simple framework

Figure 1 helps illustrate what we mean by scope, and why it’s easy to lose track of the issue of scope when we are designing or evaluating policies.

Figure 1. Factors that influence scope



Sometimes we start with a policy-relevant outcome we are trying to change, Y^* , such as wealth or obesity or high school graduation or violent-crime victimizations, which is the sum of individual-level outcomes across all people within the population, $Y^* = \sum_{i=1}^N Y_i$. We then try to identify a specific decision people make, D_i , that affects the outcome and might be responsive to some intervention, P . If the outcome is obesity, for instance, we might identify a decision related to eating or exercise.⁴

Other times, we start with the decision that we think we can change, D_i , and then think about what outcome(s) it might affect. For example, if we can encourage youth to attend school more consistently, we might hope this could not only improve schooling outcomes but, by occupying youth for large parts of every day, keep them out of trouble as well.⁵

The key point is that we usually focus on identifying the link between a single, manipulatable decision D_i and a policy-relevant outcome Y_i . This focus on individual-level treatment effects – even if we later consider how scaling the policy to reach more people might generate an

⁴ The relationship between these decisions and the outcomes may vary across people, for example because of differences in metabolism. The relationship between the decisions and the outcome may also be non-linear and change with the value of the outcome, if for example metabolism itself changes as people begin losing weight.

⁵ See for example, Jacob and Lefgren (2003).

aggregate of individual-level effects – can lead us to lose sight of some key questions we might have asked ourselves during the project-selection stage about the potential scope of our candidate interventions.

B. Share of Decisions Affected

The first question we should ask ourselves in trying to assess the potential scope of a candidate intervention is: What share of all the decisions an individual makes that affect some outcome might be affected by our intervention? Some key circumstances that give rise to larger decision coverage, and hence (all else equal) larger scope, include:

- Interventions for outcomes that depend largely on a *single one-time decision* (just a single D_i influencing the outcome Y_i).⁶ Retirement savings is a canonical example, which often critically depend upon a one-time enrollment decision like signing up for an employer-sponsored 401k plan. Another example is organ donation, which depends on one-time enrollment decisions made at key moments like the DMV for license renewal (Johnson and Goldstein, 2004), or plan-making interventions that get people to go out and vote (Gerber and Rogers, 2009 and Nickerson and Rogers, 2010). A different type of example involves the decision to purchase a capital good that lasts for some time, as when Australia encouraged people during a drought to install low-flow shower heads or water tanks for gardens (Walton and Hume, 2011).
- Interventions that target a *series of decisions made in the same context*. This makes it easier to modify a large number of decisions through changes to a small number of types of choice environments. For example, checklists for surgeons have been shown to generate large reductions in surgery-related infections, injuries and even deaths (Treadwell et al., 2014). Another example includes providing doctors with information about a patient’s overdose, which has been shown to reduce opioid prescribing (Doctor et al., 2018), or providing doctors with prescribing practices of other doctors (Scarney et al., 2018).
- Interventions that target decisions where there is a high level of *habit-formation*. For example, implementation of intentions to exercise (“what days and times will you plan to exercise next week?”) could lead to lasting behavior change if exercise is habit-forming (Milne et al., 2002 and Gollwitzer, 1999).
- Interventions that target multiple decisions that are all motivated by a shared underlying reason (that is, a single R_i links to multiple D_i in Figure 1). For example, suppose that a wide range of diet decisions – such as, choosing between meat and vegetables, or sugar and fruits, or how much fats and carbohydrates to eat, or whether or not to snack late at night – all depend on a single reason: the person’s mis-understanding of the magnitude of the health consequences of diet. Then a single intervention focused on improving this understanding could change a wide range of health-relevant decisions.

C. Importance of Decisions for Outcome

⁶ There is a separate point, different from decision coverage, about how complicated versus straightforward the one-time decision is; as Ben Castleman pointed out to us, for instance, filling out the FAFSA college financial-aid form is one-time but extremely complicated for most people.

A second question to ask for assessing scope is: How important are the decisions affected by our intervention for the outcome we care about? The impact on an outcome is not simply a mechanical function of just the number of decisions affected by some intervention: not all decisions are equally important for the behavior we are trying to change. The size of the change in a given person's outcome might depend on the share of that person's decisions that are affected, weighted by each decision's effect on the outcome for that person. This may vary across people, so for considering scope it is useful to consider different types of people – particularly the question of whether the intervention has the potential for sizable scope for *anyone*.

Luckily we often know something about the outcome itself that lets us understand the relative importance of the decisions we are trying to change.

For example, imagine trying to use behavioral science to address an important policy problem in America: gun violence. One commonly-discussed contributing factor to this problem is playing violent video games. While good research on this link is limited, the standardized effect size in an observational regression of video game playing and teen aggressive behavior has a 95% confidence interval that ranges from just -0.08 to +0.10, which likely overstates the true causal relationship because those who choose to voluntarily play lots of violent video games are plausibly predisposed towards aggressive behavior already (Anderson, 2004 and Przybylski and Weinstein, 2019).⁷ Even a policy intervention that substantially changed the playing of violent video games would be seem to be very unlikely to have a large impact on gun violence.

Contrast that with a behavioral intervention that tried to address the problem of under-saving by American households by bolstering retirement savings specifically. This is clearly an important problem. Fifty years ago, the average American family saved about 10% of their income; today the figure is closer to 2% (Sullivan, 2018). In 2016, the median household had a total of \$189,000 worth of assets, of which fully \$60,000 (32%) was in the form of retirement accounts (United States, 2017).⁸ Given that so much of household savings are in retirement accounts, increasing participation in employer-sponsored retirement plans could indeed have sizable impacts on total overall savings.

III. PRACTICAL LESSONS

A. Increasing scope can sometimes compromise scale

One way to expand the effective set of decisions changed by a candidate intervention is to focus on interventions that target a single underlying reason that shapes multiple decisions an individual makes. But that effort can sometimes come at the cost of compromising scale.

Let us return to the question of how to reduce the risk that young people are involved in gun violence. The nationwide protests that occurred in the spring of 2020 in the aftermath of the

⁷ Other studies can find larger associations with intermediate outcomes that are easier to measure in laboratory studies like aggressive affect, aggressive cognition or cardiovascular arousal (see for example, Anderson [2004]).

murder of George Floyd have led to renewed interest in ways of keeping the public safe without relying on law enforcement. One component of such a policy shift might involve increased emphasis on social programs that try to change multiple decisions that a person might make – what Hertwig and Grune-Yanoff (2017) call “boosts”.

For example, the behaviorally-informed Becoming a Man (BAM) program, developed by Chicago-area non-profit Youth Guidance, tries to help youth identify common cognitive errors they might make automatically without realizing it in high-stakes settings. One of the key lessons from behavioral science is that because deliberate cognition is effortful (what psychologists call “system 2”), we rely as much as possible on automatic responses that we have developed to be adaptive for dealing with commonly encountered situations (“system 1”). But those normally adaptive responses can sometimes get us into trouble if they get over-expressed and deployed in the wrong setting. While this general tendency is common to all people, unfortunately the consequences can be much more severe for young people growing up in neighborhoods of concentrated disadvantage where guns, drugs, gangs and crime are more prevalent.

For example, when a teenager grows up in an environment where his well-being depends on not being seen as a pushover, then it is adaptive for him to learn to retaliate when someone insults him. In distressed, high-crime areas where sources of formal social control, like police or schools, are overwhelmed, people are on their own. A person in this circumstance might learn to automatically push back when threatened to avoid additional victimizations in the future (see for example Anderson, 1999, or Papachristos, 2009). But if he is holding a gun when confronted, then (unless he really faces a serious threat to his safety) it may be better to override this heuristic. Or, if for example a teenager hears a teacher say “sit down and be quiet”, he may instinctively or automatically perceive this as a situation where he should retaliate and act up to preserve his reputation, although in this case it can lead to suspension or even expulsion from school with long-term adverse educational consequences.

Learning about these principles might affect how youth approach many decisions, particularly those involving social conflict.⁹ As is the case with many boosts, BAM hires people to coach or work intensively with participants on decision making. And these programs, which typically involve 15 or 20 contact hours with participants (sometimes more), often cover a range of different decision-making scenarios.

In Heller et al., we reported on two large-scale randomized controlled trials (RCT) of BAM with several thousand youth that showed very large declines in violent-crime arrests of 45-50% during the program period, as well as sustained gains in high school graduation rates of 12-19% of the control mean (Heller et al., 2017). The intervention is clearly solving the scope problem, given these large impacts. (Nor are these types of large behavioral impacts limited to just BAM.)¹⁰

⁹ These efforts to change general decision-making processes are in contrast to efforts to change-the-chooser by educating them about a specific decision, such as the health and other consequences of binge drinking (having more than five drinks at a time for men or four or more at a time for women).

¹⁰ In Heller et al. (2017), we also report on the results of a CBT-type intervention that was randomly assigned within the Cook County, Illinois Juvenile Temporary Detention Center (JTDC), which reduced readmission rates through 18 months following JTDC exit by around 16 percentage points (21%). In the same spirit, Blattman et al. (2017) study a CBT program in Liberia that worked with men between the ages of 18 and 35 and found a reduction in anti-social behaviors by 0.2 standard deviations. More recently, other members of the University of Chicago Crime Lab

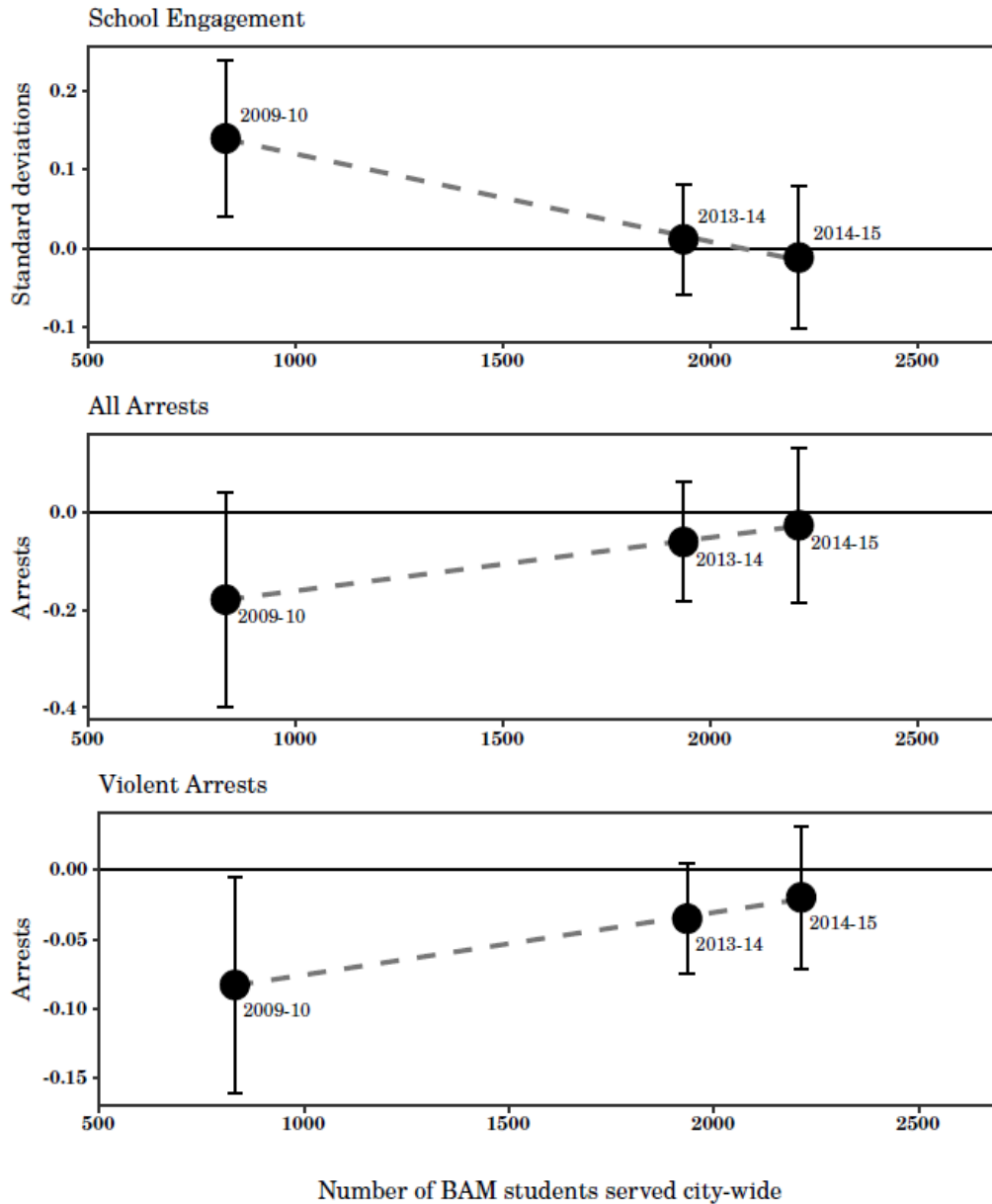
But, as access to the program expanded, the average impact may well have declined. This can be seen in Figure 2, where we plot the number of youth participating in BAM citywide in a given academic year on the x-axis against the average effect of BAM on program participants who were in any of our RCTs in that year.¹¹ In the later years with substantially larger program participation levels, we can no longer reject the null hypothesis that impacts equal zero, although given our statistical sample sizes here, we also cannot reject the null hypothesis that the program effects are constant at different scales. So, we would consider this suggestive but not conclusive evidence that effectiveness may have declined as scale increased. While Youth Guidance has made some changes to the program intended to enhance fidelity and support program quality at scale which might strengthen its effects, we do not at this point have any evidence on how those changes have moderated effectiveness.¹²

have been studying the Choose to Change (C2C) program for youth at elevated risk for violence involvement and find the program reduces their risk of violent-crime arrest by around 50%. See Abdul-Razzak and Hallberg (2018), the results of which were discussed in Kirp, David. “How to end the cycle of violence in Chicago.” *The New York Times*, September 13, 2018.

¹¹ Note that not every youth receiving BAM in a given year was in one of our RCTs, but nonetheless from a program-operations perspective at least the total number of participants citywide in a year is the relevant parameter.

¹² For example Youth Guidance has made significant investments in training and coaching on counselor competencies and program delivery, and has worked with National Implementation Research Network’s (NIRN) implementation science framework to strengthen consistency and replicability over sites and over time. As of yet there have been no further studies to establish evidence about the impacts of these adjustments.

Figure 2. BAM impact by scale of BAM delivery city-wide



Note: TOT point estimates are displayed as a function of the number of BAM students served city-wide in a given year. Point estimates measure outcomes among study participants while the number served city-wide includes all BAM participants, including those outside of the study sample. School engagement outcomes are measured in standard deviations while arrest outcomes are measured in arrest counts. Participation for the IV is measured as attending at least one session during the year. The estimates for 2009-10 include participants from Study 1 only; the estimates for 2013-14 include participants from Study 2 and Study 3; the estimates for 2014-15 include participants from Study 2, Study 3, and Study 4. Dashed lines reflect the linear best fit. Standard baseline covariates and randomization block fixed effects are included in each model and standard errors are clustered at the student-level.

BAM has a number of shared features with the sort of ‘boosts’ or social programs that are often used to address social problems that can help solve the scope problem but may as a result exacerbate challenges of scale. For example:¹³

- The program is not mechanical; being a good counselor clearly requires a specific skillset. People with those skills who are willing to work at a given wage are presumably in finite supply, so to recruit additional skilled counselors as the program expands, higher wages would presumably be required. Yet, as the program expanded in Chicago, wages (as far as we know) did not substantially increase, so it is possible average counselor effectiveness may have declined (even though the crude proxies for effectiveness we can measure did not seem to change very much over time).
- While BAM has a curriculum, by its nature it involves some discretion on the part of individual providers about how it is delivered session-to-session and how providers tailor specific actions to build relationships with specific participants. This makes it difficult to fully write down the “formula” or recipe for how the program works (indeed Youth Guidance’s own internal implementation measures are weakly correlated with one another, and in some cases even *negatively* correlated).

What is the lesson for project selection? The degree to which a candidate intervention involves a production process that can be fully described will sometimes require some careful on-the-ground observation. The degree to which the intervention relies on inputs that will be harder to obtain as scale increases can be inferred from some combination of introspection on whether the intervention’s inputs seem specialized or not, whether there are other interventions out there that rely on similar inputs whose experiences can be studied and learned from, and in some cases, pilot-testing combined with what Davis et al. termed “scale up experiments” (Davis et al., 2017).

B. Scope and scale challenges as marginal cost questions

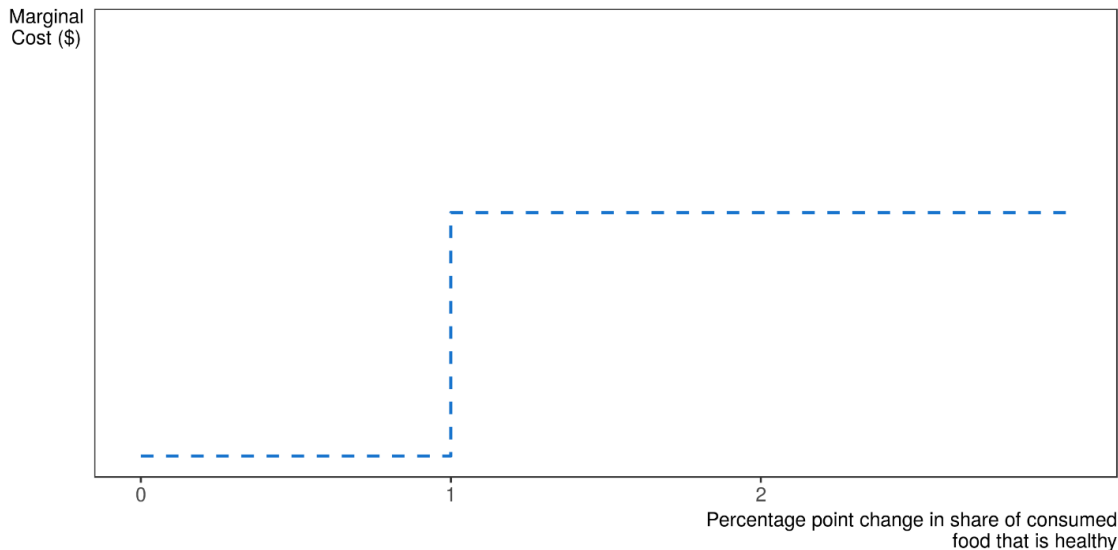
As noted above, a key challenge in improving children’s healthy eating by swapping the location of fruit and cafeteria in the school cafeteria comes from the fact that children only eat a modest share of all their meals each week in a school cafeteria. But of course, that’s not the *only* setting in which children eat meals where we might swap the locations of healthy and unhealthy food. What are the practical implications for intervention to have to fall back on some other setting? How would we even begin to think about this?

The simple idea from economics of marginal costs gives us an easy way to think about this. Let the x-axis in Figure 3 be the change in the share of consumed food that is healthy; the y-axis is the marginal cost of achieving a given change. Swapping fruit and candy in the school cafeteria provides a very low-cost way to increase the share of food consumed by children that is healthy by about 1 percentage point (per our calculation above from the results presented in Hanks et al.,

¹³ As discussed in the technical appendix, changes in the characteristics of program participants and school- or community-level contexts as the program expands are a partial explanation for why the average effectiveness BAM seems to have declined as the scale of the program- that is, the number, locales, and types of students and schools served - increased. These findings are consistent with some heterogeneity in how youth respond to the intervention, but this is just a partial explanation.

(2012). To achieve larger changes in healthy eating requires changing what they eat in other settings, where it becomes more difficult (hence costly) to change their eating – represented by the increase in the marginal cost curve for the 2nd percentage point change in healthy eating, which corresponds to the next-lowest-cost way of changing their eating.

Figure 3: Marginal costs of changing healthy eating among children (hypothetical example)



Of course, it will not always be possible to precisely trace out the marginal cost schedule like in Figure 3 for a given intervention, but even rough attempts to approximate it can help researchers and policymakers alike begin to consider questions during the project selection stage like:

- Over how large a range does the marginal cost curve stay at a low level?
- How quickly does it begin to increase thereafter – that is, by how much does the cost of expanding the scope of the intervention begin to increase?

The idea of marginal costs can also be useful in helping researchers and policymakers construct an optimal portfolio of interventions to address some big problem in situations where there is no single large-scope intervention to be found. Suppose the first percentage point of healthy eating for a given person changed through rearranging food in the school cafeteria costs (say) 50 cents, while the first percentage point of change in healthy eating due to an intervention that provides nutritional information costs 75 cents. We should obviously do 1 percentage point worth of cafeteria food substitution for starters. If the cost of changing the next percentage point through cafeteria food rearranging is less than 75 cents we should do that as well, but if it's more than 75 cents we should combine 1 point of cafeteria re-arranging with 1 point's worth of nutritional education as well.

Taking seriously the idea of trying to achieve large changes in outcomes can substantially change our thinking about what intervention, or mix of interventions, is most promising to study and/or should eventually be adopted. In our job as scientists carrying out studies, we typically focus on

identifying interventions where the marginal costs of reaching the initial subset of participants, or subsets of relevant decisions, are low. That is, if we think about this curve as relating marginal cost (y-axis) to number of people or share of relevant behavioral decisions affected (x-axis), our search for promising interventions to pilot-test focuses on the *intercept* of this curve (y-axis location at x-axis value of zero). But for purposes of achieving large, population-level impacts on social conditions, we need to be focusing much more on the marginal cost curve's *slope* over a large range of the x-axis as well.

C. Scale and scope as interdependent challenges

While scale and scope are conceptually distinct, they interact in important ways. For example, suppose our intervention involves swapping the location of the fruit and candy at the check-out line of the movie theatre. Scope involves asking the question: For each person (or "type" of person), how much of the eating decisions that impact health could be affected by this intervention? Suppose the data revealed that no one goes to the movies more than once per month. In this case, considering scope alone is enough to tell us that there is no way for this intervention to substantially reduce obesity rates.

But suppose the data revealed that some people go to the movies three times a day and do all of their eating there. In that case, understanding the potential aggregate-level impact of the intervention requires considering the follow-up scale question: What share of people in the general population are of the type that go to the movies all the time versus the type that goes rarely? In that sense, scope questions are asymmetrically informative; they can by themselves help rule out a candidate intervention's potential for large-scale population-level impact, but by themselves cannot establish that potential without some additional consideration of scale.

IV. CONCLUSION

We have argued that the increased involvement of social scientists to help make big progress on major social problems will benefit from paying more attention to scope – that is, how to affect a large share of the decisions that shape a given person's behavior. How would this change what social scientists do in practice?

- Many RCTs are about developing basic understanding of human behavior, rather than identifying solutions to policy problems. That useful enterprise should proceed as is.
- Some social scientists may be mostly interested in identifying interventions that pass a benefit-cost test, regardless of aggregate impact on a problem. That useful work should also proceed.
- For those interested in achieving large changes in outcomes, assessing scope helps determine when we can and cannot get to big impact by assembling a portfolio of small-impact policies. Assessing scope can help us better understand when this is not possible because:
 - the marginal cost of changing some decisions that are necessary to affect the ultimate social outcome is high
 - a very large number of interventions would be required and so we might run up against constraints on R&D bandwidth or the capacity and patience for innovation by external partners

- even the accumulation of interventions might together have limited scope, so we can in those cases also assess the public’s willingness to support multiple efforts to solve a given problem in order to avoid, or at least understand the potential for, “crowd out” across policies (Hagmann et al., 2019).
- In cases where we cannot aggregate up to large impacts through a portfolio of small-impact policies, we might prioritize work on interventions where scope challenges are less pronounced: those that target outcomes that depend largely on a *single one-time decision*, or on a *series of decisions made in the same context*, decisions where there is a high level of *habit-formation*, or a *single reason that lies behind multiple decisions that are relevant for the outcome*.

Operationalizing the framework we lay out here is not trivial in practice, but we hope it is a useful start in prioritizing efforts within social science to promote social progress.

REFERENCES

- Abdul-Razzak, N. & Hallberg, K. (2018)
<https://www.socialscienceregistry.org/trials/933/history/33957>.
- About Us. (n.d.). Retrieved from <https://www.povertyactionlab.org/about-j-pal>.
- Al-Ubaydli, O., List, J.A., LoRoe, D., & Suskind, D.L. (2017) Scaling for economists: Lessons from the non-adherence problem in the medical literature. *J. Econ. Perspect.* **31**, 125-144.
- Anderson, C.A. (2004) An update on the effects of playing violent video games. *J. Adolesc.* **27**, 113-122.
- Angrist, J., Azoulay, P., Ellison, G., Hill, R., & Feng Lu, S. (2017) Economic research evolves: Fields and styles. *American Economic Review, Papers and Proceedings*, *107*(5), 293-97.
- Angrist, J.D., & Pischke, J. (2010) The credibility revolution in empirical economics: How better research design is taking the con out of econometrics. *Journal of Economic Perspective*, *24*(2), 3-30.
- Banerjee, A., Banerji, R., Berry, J., Duflo, E., Kannan, H., Mukerji, S., Shotland, M., & Walton, M. (2017) From proof of concept to scalable policies: Challenges and solutions, with an application. *J. Econ. Perspect.* *31*, 73-102.
- Blattman, C., Jamison, J.C., & Sheridan, M. (2017) Reducing crime and violence: Experimental evidence from cognitive behavioral therapy in Liberia. *Am. Econ. Rev.* **107**, 1165-1206.
- de Rezende, L.F., & Eluf, J.N. (2016). Population attributable fraction: Planning of diseases prevention actions in Brazil. *Rev Saude Publica*, *50*(30). Retrieved from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4902656/>.
- Davis, J.M.V., Guryan, J., Hallberg, K., & Ludwig, J. (2017) The economics of scale-up. *NBER*, **23925**.
- Dinner et al. (2011). *Journal of Experimental Psychology: Applied*, *17*(4), 432.
<https://doi.org/10.1037/a0026470>.
- Doctor, J.N., Nguyen, A., Lev, R., Lucas, J., Knight, T., Zhao, H., & Menchine, M. (2018) Opioid prescribing decreases after learning of a patient's fatal overdose. *Sci.*, *361*, 588-590.
- Gerber, A.S., & Rogers, T. (2009) Descriptive social norms and motivation to vote: Everybody's voting and so should you. *J. Politics*, *71*, 178-191.
- Gollwitzer, P.M. (1999) Implementation intentions: Strong effects of simple plans. *Am. Psychol.*, **54**, 493-503.

- Hagmann, D., Ho, E.H., & Loewenstein, G. (2019) Nudging out support for a carbon tax. *Nat. Clim. Change*, 9, 484-489.
- Hamermesh, D.S. (2013). Six decades of top economics publishing: Who and how? *Journal of Economic Literature* 51(1), 162-72.
- Hanks, A.S., Just, D.R., Smith, L.E., & Wansink, B. (2012) Healthy convenience: nudging students toward healthier choices in the lunchroom. *J. Pub. Health.*, 34, 370-6.
- Heller, S.B., Shah, A.K., Guryan, J., Ludwig, J., Mullainathan, S. & Pollack, H.A. (2017) Thinking, fast and slow? Some field experiments to reduce crime and dropout in Chicago.” *Q. J. Econ.*, 132, 1-54.
- Hertwig, R., & Grune-Yanoff, R.T. (2017) Nudging and boosting: Steering or empowering good decisions. *Perspect. Psychol. Sci.*, 12, 973-986.
- Jacob, B.A., & Lefgren, L. (2003) Are idle hands the Devil’s workshop? Incapacitation, concentration and juvenile crime. *American Economic Review*, 93(5), 1560-77.
- Johnson, E.J., & Goldstein, D.G. (2004) Defaults and donation decisions. *Transplantation*. 78, 1713-16.
- Kirp, D. (2018, Sept. 13) “How to end the cycle of violence in Chicago.” *The New York Times*.
- Milne, S., Orbell, S., & Sheeran, P. (2002) Combining motivational and volitional interventions to promote exercise participation: Protection motivation theory and implementation intentions. *Br. J. Health Psychol.* 7, 163-184.
- Muralidharan, K., & Niehaus, P., (2017) Experimentation at Scale. *J. Econ. Perspect.*, 31, 103-24.
- Nickerson, D.W. & Rogers, T. (2010) Do you have a voting plan?: Implementation intentions, voter turnout, and organic plan making. *Psychol. Sci.*, 21, 194-199.
- Przybylski, A.K., & Weinstein, N. (2019) Violent video game engagement is not associated with adolescents’ aggressive behaviour: evidence from a registered report. *R. Soc. Open Sci.* 6, 171474.
- Sacarny, A., Barnett, M.L., Le, J., Tetkoski, F., Yokum, D., & Agrawal, S. (2018) Effect of Peer Comparison Letters for High-Volume Primary Care Prescribers of Quetiapine in Older and Disabled Adults: A Randomized Clinical Trial.” *JAMA Psychol.*, 75, 1003–1011.
- Social Science and Contemporary Social Problems. (2019, August 21). Retrieved from <https://items.ssrc.org/from-our-archives/social-science-and-contemporary-social-problems/>

- Sullivan, B. (2018, August 28). Once again, Americans are not saving enough. *Market Watch*. Retrieved from <https://www.marketwatch.com/story/once-again-americans-are-not-saving-enough-2018-08-28>
- Treadwell, J.R, Lucas,S., & Tsou, A.Y. (2014) Surgical checklists: a systemic review of impacts and implementation. *BMJ Qual Saf.*, 23, 299-318.
- Walton, A. & Hume, M. (2011) Creating positive habits in water conservation: The case of the Queensland Water Commission and the Target 140 Campaign. *Int. J. Nonprofit Volunt.*, Sect. Mark. 16, 215-224.
- United States. Federal Reserve Board. (2017, September). *Changes in U.S, Family Finances from 2013 to 2016: Evidence from the Survey of Consumer Finances*. Retrieved from the Federal Reserve Web site: <https://www.federalreserve.gov/publications/files/scf17.pdf>.