Original Research Article

Dubious Evidence, Valuable Information

Evaluation Review
2025, Vol. 0(0) 1–15

DOI: 10.1177/0193841X251324594
journals.sagepub.com/home/erx
Sage

Eugene Bardach[1]

Abstract

I explore how "low-quality evidence" from program performance might still be useful in decision-making. Conceptually, a local government named "Here" is motivated to consider a program from "Elsewhere" that seems to show year-over-year exemplary performance. Here must manage five sources of uncertainty about whether and how to extrapolate from Elsewhere: chance in assessing Elsewhere's performance; illusion due to confounding variables; estimating the several powers of the program's components; substitutions in the design process made by Elsewhere and contemplated by Here; and estimating whether in the final analysis Here can meet its own breakeven criterion for going ahead. Here can begin with Elsewhere's experience, but it still must do much thinking and information-collecting on its own.

Keywords

extrapolation, implementation

Policy is often a battleground where hopefulness faces off against skepticism. I want to explore one shadowed corner of that battleground, where skirmishes over a single, but spectacularly successful, example play out. This example

[1]Goldman School of Public Policy, University of California, Berkeley, CA, USA

Paper based on Peter H. Rossi award presentation at the 2024 Research Conference of the Association for Public Policy Analysis and Management

Corresponding Author:
Eugene Bardach, Goldman School of Public Policy, University of California, 2607 Hearst, Berkeley, CA 94720, USA.
Email: ebardach@berkeley.edu

comes from one city, "Elsewhere," which another city, "Here," is considering as a model for its own version of the policy. "Here" is moved by hopefulness, although with some local skeptics being dragged along. These local skeptics are joined by outsider skeptics from the scientific policy evaluation community who see only one data point, and one that is not even randomly selected at that. They also are conscious of Peter Rossi's "Iron Law" of policy evaluation, which holds that actual policy impacts trail off towards zero when the policy undergoes successive evaluations.[1] Even though Rossi qualified this law once it became dogma, it is probably true enough to make scientific evaluators into covert, if not always overt, skeptics.

The skirmishes between the party of the Hopefuls and the party of the Skeptics are largely over how to handle uncertainty. This uncertainty attaches to beliefs about what is the case, what has been the case, and what will probably be the case. In order to draw the sharpest outlines of these skirmishes, let us assume that the available data are extremely limited: Elsewhere's apparent exemplary performance is based purely on data comparing short time periods before and after the policy intervention.

Our discussion will address five areas of uncertainty that appear in most policy debates. These five begin with Elsewhere's performance data and continue to other uncertainties involving both Elsewhere and Here.

Chance. Might Elsewhere's spectacular success be merely a product of chance?

Confounding. If it is not a product of chance, might it be a product of other features of Elsewhere that just happen to be correlated with the policy introduction?

Power of the components. Of all the components of the policy, which might be the most powerful, and how powerful might it be in Elsewhere and, also, in Here?

Substitution. Since Here must substitute design elements for those in Elsewhere, how might this affect policy performance?

Decision. What might be the odds that Here's policy outcomes would show benefits exceeding costs (both very broadly construed)? Are the odds good enough?

Framing and Format

Framing. Uncertainties are numerous and so are conclusions about them. It would be useful to have a unifying frame. Start with the central question asked by Here: "If we extrapolate from Elsewhere, how likely is that the outcomes meet our "breakeven" criterion." The breakeven criterion is familiar from benefit-cost analysis. No matter how Here chooses to define "benefits" and "costs," or to weight and aggregate them, Here can draw a breakeven boundary along which the expected net benefit of a policy changes from

positive to negative or vice versa with a move in any direction away from the boundary. To be worthy of adoption, any policy choice must fall on the positive (net-benefits) side of the boundary. For example, if Here contemplates building another wastewater treatment plan and it costs $3 million, the expected benefits must be at least $3 million before a positive decision is warranted.

Format. I use an unconventional format for the following discussion. It is a dialogue. For each of the five discussion areas, I begin with a Hopefulness take on the uncertainty. This is followed by the Skeptic's view. It ends with a Realist synthesis.

I necessarily speak for the Realist here. If not I, who else? Actually, "Realist" has no firm meaning where managing uncertainty is concerned and where even a Bayesian would not know quite what to make of the low-quality "evidence" available. I believe I lean, just a little, towards hopefulness. As I am writing for a journal audience made up primarily of evaluation professionals, and who therefore are disposed towards skepticism, I aim to offset this disposition, even if just a little.

Analogy with "evidence" in law. I hope to borrow a little bit of legitimacy for this effort from the inspiring works of law professor Frederick Schauer (Schauer, 2022). His lifelong specialty has been how the legal system handles "evidence" in all its dimensions. In a phrase, his view is: prefer utility over scientific accuracy. Do not throw out useful evidence just because it does not meet a high standard of scientific accuracy. Depending on the stakes and other conditions, it might, or might not, be useful. We convict criminals for offenses if proof is "beyond a reasonable doubt," but award civil damages if the plaintiff's case is supported only by "a preponderance of the evidence."

Evidence and information. A first step towards a clarifying dialogue is semantic. I substitute the term "information" for "evidence." The domain of "evidence" is too narrow with respect to the domain of possible consequences (prison/no prison; adopt/don't adopt…) and too hard edged (meets some standard /doesn't meet some standard). "Information," like "evidence," begins with some observed datum but then can be used for many different purposes by many different minds and by many different institutions. Information can sometimes serve as evidence, but not always. In our assumed domain of Hopefulness/Skepticism skirmishes, the information in question can be used to inspire, warn, guide, search, and enlighten. It can also be used to legitimate or delegitimate.[2]

Tiny Homes

To help ground the following abstract discussion of uncertainties in the Elsewhere-to-Here extrapolation process, I will use as a source of examples a stylized real-world case, a program to deal with homelessness generally

known as Tiny Homes.[3] What I know of it is mainly based on a lengthy essay that appeared in the San Francisco *Chronicle*[4] of October 13, 2024, written by Elizabeth Funk, the CEO of an NGO, DignityMoves, that promotes this program in various cities, including San Francisco, and a subsequent interview with her and with a Goldman School alumnus who is particularly well informed about housing for the homeless.

The policy idea behind Tiny Homes is to provide each homeless person or family a small, stand-alone, prefabricated cabin with a lockable door. It is to be situated on unused public land or on leased private land, land which could be reclaimed on short notice. Such interim housing could be done much, much more rapidly and inexpensively than providing permanent housing. Perhaps the earliest pioneer in Tiny Homes is San Jose. In one year, while San Jose's homeless population decreased by some 10%, in the same time period California's increased by the same amount. About three years ago San Francisco was hoping to imitate San Jose's success. As of November 2024, the DignityMoves web site carries this description of the program in San Francisco:

> A partnership with Tipping Point Community [a local philanthropy], The Department of Homelessness and Supportive Housing, Urban Alchemy and Home First Services, this pilot project is the first of its kind in San Francisco. Located at 33 Gough Street in the South of Market neighborhood, the site provides 70 private, dignified rooms for individuals experiencing homelessness. Each room has a bed, a desk and chair, heating, a window, and most importantly a door that locks. The site has been thoughtfully designed by Gensler and PAE Engineers to create a welcoming community and foster a sense of belonging. The community also includes case manager offices, extensive dining and community spaces, a computer lab, pet area, community gardens, and ample storage for residents' belongings.

Five Uncertainties

Chance

Hopefulness. Elsewhere is inspiring. Whatever the risks, they are worth taking.

Skepticism. Because Elsewhere is not only exemplary, but spectacularly so, that is exactly why it is inspiring. But the downside here is that its extremely good performance is almost surely in some measure due to chance and therefore in some sense illusory.

Realism. This problem can be offset in some degree by looking at trends on both the before and after sides of Elsewhere's intervention. Trend data smooth out random fluctuations. Unfortunately, as we move further towards the extremes of any chance-affected distribution of observations, the odds increase that any particular observation has a large chance component. In many

cases these trend data within Here can be usefully compared to trend data in some setting that, we have reason to believe, is not affected by the focal intervention.[5]

Illusion

Hopefulness. Just look at the difference in rates of homelessness before and after Tiny Homes, and all this at a time when the rates of homelessness were actually going up outside of Elsewhere.

Skepticism. Yes, there is a correlation. But that does not mean it is actually causal. Other factors that go along with Tiny Homes might be the real causes of the decline in homelessness. As Tiny Homes became available, local police might have decided to increase pressure to disband tent and other encampments. Or perhaps the local economy shifted dramatically for the better and permitted extra revenue which got spent on more and better case management. Or business investment in San Jose brought new jobs and therefore new employment opportunities and income to the homeless.

Another source of appropriate skepticism about Elsewhere's replicability is that it might be blessed by unusual access to high-performing resources, like a few very well-placed and talented administrators or a very solid revenue base. Or perhaps it is the absence of certain legal, institutional, and political constraints that have permitted modular and below-code construction which would have been done by for-profit housing developers anyway.

Realism. A possible offset to this "confounding variables" problem is probing to look at what these variables have done in Elsewhere with respect to similar policies. Similarly, what have they done outside of Elsewhere? Such information is hardly conclusive but is suggestive nonetheless. If real-life comparisons are not available, try thought experiments. What if the possibly confounding variable had not been present, might the observed outcome have occurred anyway? What if the program had not been in place, would the confounding variable have been sufficient to have caused the observed outcome?

Power

The Tiny Homes program, like most public programs today, has three component subsystems that combine to produce performance: the energy-providing driver mechanism; the value-creating production mechanism; and the implementation-supporting mechanism. "Performance" typically means ameliorating some condition that the relevant political community defines as a problem that should be addressed by its political jurisdiction.

I simply assert these three components without detailed justification. However, they are the product of a model of policy programs in general that springs from efforts at reverse engineering.[6]

It is the energy-providing component, the "drivers," that powers the production mechanism and therefore, indirectly, the whole program. Drivers typically draw on energy sources that are free or very low-cost given the substantial work that they do within the system. Hydrocarbon bonds, full of energy created by nature, have their energy extracted and channeled by automotive mechanisms that can then move large masses of steel and glass. Photosynthesis harvests sunlight and converts the energy to acorns and oak leaves. In the human sphere, we commonly find drivers in self-interested human nature to fashion markets that then increase value through bargaining, searching, and exchanging. One driver in the Tiny Homes program is the natural desire of human beings to improve their feelings of security, shelter, and privacy. Another such driver is the natural desire to buy a functionally equivalent product (prefab housing) for fewer resources.

The production mechanism is the most tangible and visible component of the program machine. In the Tiny Homes case, the production mechanism includes modular housing units, a procession of homeless individuals moving into and out of them, the funds for construction of the units and the contractors' work that went into them, the sanitation department apparatus that is part of the ongoing maintenance operation. Although the production mechanism is driven by the energy-providing component of the system, it has its own internal machinery that is, in a design sense, largely independent and has its own ways to strengthen or diminish the energy supplied by the drivers. In the world of physical nature, it is the energy in a coiled spring that drives the hands of a wall clock, but it is the cogs and wheels of the production mechanism that extract that energy and convert into work and the social value the work creates. In the Tiny Homes case, the nature and degree of security provided by its consciously designed lockable doors, might greatly augment or diminish the nature-embedded willingness of homeless persons to move there.

The implementation mechanism is basically the administrative and political apparatus that supports (and sometimes disrupts) the production component. It contains, for instance, the agency that manages procurement contracting for private-sector work but also the detailed and sometimes dysfunctional procedures that govern its operations and procedures.[7] It also contains all the political forces that defend the program and those that abuse its resources and pervert its goals.

Hopefulness. Both drivers in the Tiny Homes case, mentioned above, are self-evidently rich and universal energy sources. Both drivers seem powerful and uniform, varying but little from person to person or site to site.

Skepticism. However, it is not enough to know what the drivers are; one still needs to estimate how powerful they are in general and, even more to the point, how powerful they would prove to be in Here.

Information from Elsewhere is suggestive but just barely. The power of the drivers does not appear alone. The production component sometimes magnifies and sometimes diminishes the power of the drivers. In the Tiny Homes case, the quality of on-site case-management services affects the capacity, and probably the desire, of homeless individuals to move from a familiar to an unfamiliar living site. The implementation component also affects the city's ability to supply the modular housing units expeditiously and cost-effectively. It is hard for Here, just by knowing a lot about Elsewhere, to gauge how powerful the drivers are—or how the other components affected them.

Realism. Here can consider each component separately, beginning with the drivers, and then move through the others, either alone or in combination. Thought experiments can be helpful: "What if … were different in this or that way, would the power of the drivers have been increased or decreased?" Here can look for information outside Elsewhere as well: "What happens in general when public sector procurement processes get rolling?"[8] Information and ideas gleaned thereby can help Here select and interview helpful informants in Elsewhere. Staff in Here can also find appropriate interviewees in Here.

At this point decision-makers and staff in Here must confront the problem of "context."[9] The professional evaluation community is well aware of the problem that "context" affects the power of the program as a whole and its several components to function well or poorly. Often heard in Here is some variation of: "Yes, that's fine for Elsewhere, but we [are much poorer, are experiencing a growth spurt, have a very different demographic picture, have a month of below-freezing days, have an ancient infrastructure, have a long tradition of physical risk-taking and self-reliance…]"[10] The important contextual features can also be much more detailed. In San Jose, the political momentum behind Tiny Homes included a desire to stop a source of sewage inflow into a local stream. Local officials only slightly concerned with problem-solving but more concerned with avoiding conflict or change at all may use alleged differences in context as a cover for their inertia.

A scaled up program almost always presents a different, and less favorable, context from the pilot or demonstration program from which it might be extrapolated. It normally requires new infrastructure investments, above-average personnel competencies and motivations, and protection from bureaucratic competition. Although the political and bureaucratic dynamics of scaling up frictions are not clear from their report, the disappointing results from a very simple local government "nudge" program are suggestive (DellaVigna et al., 2024).

Substitutions

The choice of drivers aside, substitutions of one choice or one solution or one mechanism for another are the beating heart of the policy extrapolation

process. Goals come and go as coalition composition shifts, as compromises are struck, as political leaders turn over, and as delays and deadlines threaten momentum. But the observable elements of the Elsewhere example generally do not include this record of substitutions. They are sometimes just as relevant to Here's policy design as the elements of the present and just-past, however. What did Elsewhere discard and why? Might these choices be more desirable for Here than those that Elsewhere did in fact embrace? If so, then Here staff might consider interviewing knowledgeable individuals in Elsewhere.

Of all potential substitutions, though, the most important concern catalysts. These are the critical but seemingly small elements of a policy system that make the rest of the system function well—or, in their absence, fail. One example is some middle-level bureaucrat in Elsewhere who has a deep understanding of the homelessness issue, a workable theory as to how to make Tiny Homes work in Elsewhere, and enough leadership resources to provide for the production and implementation-supporting functions. If such catalysts are lacking in Here, one may need substitutes.

Hopefulness. But how to conceive of them and then search for them? Information from Elsewhere's success could be helpful to Here. First, a scan of the machinery in Elsewhere could reveal the nature and extent of possible catalysts. Here would need to find substitutes in its own environment. The raw materials could be rare and their development erratic. Such a search could be frustrating. It could be done a lot more efficiently if information from Elsewhere could be used to sharpen the conceptual focus of such a search and to guide search efforts towards the most productive possibilities.

Skepticism. Catalysts indeed! The concept is blurry and the agenda easily manipulated by opportunists. For what it is worth, a rigorous study of how "nudge" policies did or did not get scaled up in 73 localities, following successful RCT's, found that the participation of so-called "champions" had no effect (DellaVigna et al., 2024).[11]

Realism. Substitution is everywhere and endless. Once a program in Here is up and running, flaws will be found and improvements suggested. These too are substitutes. Extrapolation from Elsewhere occurs only at the beginning of this process. As the program evolves, it is the later incarnation of Here that needs to extrapolate from information available from its earlier incarnation in Here, and to make indicated substitutions.

Decision

In the end, Here faces a decision: do or do not attempt to extrapolate from Elsewhere? To answer this question, here is a useful principle: the drivers of performance should be stronger than the combined frictions, for example, very high prefab costs, arising from channeling and implementation. This principle would be of little use were it not for the fact that the net strengths in this case

are always contingent on context. While "drivers" always has a plus sign and "frictions" usually a minus sign, occasionally "frictions" are positive, in which case we should probably rename them "supports," for example, high prefab costs stimulating search for lower-cost sites. Thus, if Here succeeds with its own Tiny Homes program, it is likely that, by learning from Elsewhere it has managed to replicate the strengths of Elsewhere's drivers while turning the contingently negative production and implementation mechanisms into positive supports.

But how should these contingencies be aggregated and then factored into a final decision? The answers are debatable and we should welcome such a debate. Nevertheless, the Hopeful and the Skeptics have their differing opinions, which probably reflect emotions more than ratiocination.

Hopefulness. Assume that the analyses of benefits and costs are themselves realistic. Then assume that no contingencies other than the ones brought to light so far exist and that realistic estimates have been made for all of them. Then, using some version of holistic thinking—we cannot say which, but all are debatable (Schauer, 2022, pp. 31–33)—check that the breakeven criterion is satisfied. If the answer is yes, then review the (virtual) list of contingencies to find those that would flip a positive decision to a negative one. Go back to the entire list to see if any other contingencies, alone or in combination, would offset any of these.

Skepticism. Skeptics place the burden of proof on extrapolation and as a result lean towards ignoring information from a single example. The less sophisticated among the skeptics cite the incisive Yiddish saying "For instance is not a proof," as though policies, extrapolated or otherwise, did not take place in the future, which by definition rules out proof of anything whatever. Though having in mind no good methodology to substitute for the Hopefulness approach, Skeptics are moved to not favor the proposed program. The spirit behind the Yiddish saying is powerful, whether based on logic or not.

Realism. Among policy-makers, in the United States at least, avoidance of costly errors weighs heavily. Skepticism therefore has a political advantage. It also has the advantage of a scientific propensity to demand fairly high standards of evidence (.05%…) before accepting any hypothesis, including that of a proposed policy being likely to meet the breakeven criterion.

The political standard requires respect, but the scientific standard, not being relevant, is not. But what should that political standard be? The precautionary principle, often invoked over new technologies, offers a conceptual approach: let the stakes determine the standard. Given their environmental and philosophical concerns, supporters of the precautionary principle choose a very high, often close to impossible, standard of certainty. But standards can and should vary. in the law criminal guilt must meet a "beyond a reasonable doubt" standard; civil damages, a "preponderance of evidence" standard; and, for instance, civil commitment of the (supposed) mentally ill a "clear and

convincing evidence" standard. In the case of policies such as Tiny Homes, Here could choose its own standards, which would probably be based on avoiding budgetary waste. It might be "15 % more likely to work well [break even] than not." If it would have especially heavy negative impacts on a normatively protected minority group, the standard might be 25%. Minimizing possible regret over possibly catastrophic outcomes might require 90%.

Estimating whether aggregated probabilities—all the observable risks and their interactions—meet or exceed some breakeven benchmark for the success or failure of a public policy is difficult and consensual methodologies do not exist. Systemic performance probabilities are fiercely complicated, and holistic strategies, as I noted above, do not at this point suffice.

Beyond the Data Point

Here, the user, is the central actor in this drama of policy decision-making; and the evidence available from Elsewhere is, from Here's point of view, only "information." From the world's point of view, this downgrading of evidence to information is no great loss. Evidence is never really the sole basis for decision. That basis, at best, is always evidence-plus, in which the plus part refers to how the actor uses the information, including exemplary examples.

How, then, to use the exemplary example to the best effect? The "exemplary" part tells us that there is a lot of room for absorbing error and still coming out on the desired side of our breakeven boundary. As to the idea of a non-exemplary example, if it is not a source of evidence about the future, the example is still information about the past: (1) the program has appeared to work very well in at least one place, Elsewhere, and (2) that fact has, in a Bayesian fashion, caused us to update our priors about whether it might work in the future in Here. Just how much we update this prior belief depends on our reasoning about what it appears has already happened (in Elsewhere) and what we believe *will* happen in Here by way of doing further and more sophisticated interviewing and analysis.

In this construction, a useful program evaluation is a *dyad*—an Elsewhere and a Here—rather than a single data point. To be more precise, the dyad is the data point. Just how a great multiplicity of dyads might be handled by program evaluation is beyond our scope, however. But it is clear that this construction of dyad-driven evaluation brings policy evaluation, which is about the past, a little closer to policy analysis, which is about the future, and a future that is full of design choices largely in the hands of Here.

Is This Evaluation at All?

Yes, this is "evaluation," though the practitioners are users and not producers, laymen and not experts, numerous and not circumscribed, highly distributed and not concentrated, and need help, directly or indirectly, from evaluators that are more expert and credentialed. Consider the millions of customers searching the market for a new car, just as local decision-makers and analysts are searching for useful policy ideas.[12] They are all "evaluators" of a sort. They use as "evidence" information from many sources: auto makers; auto enthusiasts and their networks and publications; disinterested consumer research sources; their friends and neighbors, especially recent purchasers; themselves and their assessment of their own needs, desires, resources, and ultimately, of course, their own breakeven calculations. This is an instructive analogy. True, their "evaluations" might be imperfect and ultimately mistaken. But, like policy decisions, all their efforts at utility-maximization are simply gambles. As psychologist Daniel Gilbert observes, even though they are trying to act in what they take to be their own interests, they are "stumbling on happiness" (Gilbert, 2006).

Perhaps the policy professionals could stumble less if they had assistance from the program evaluation community. But how? What sort of assistance? It should certainly aim to help these professionals, but it should aim even more to serve the public interest. One suspects that, considering the vast number of policy extrapolation efforts carried out by local and state governments and their networks of NGO's, even modest improvements in their evaluation and policy analysis capacities would increase net social welfare by a lot. And that, after all, should be the main objective of the professional evaluation community.

But that community of program evaluation professionals might reasonably ignore this normative supposition. The community prides itself on expertise, rigor, peer appreciation, and increasing social welfare in its own way. Their efforts, they believe – and they are right – should be conserved for the projects that would benefit from their more powerful analytic methods. Moreover, low-tech methods are of little professional interest. Studies based on such methods do not make it into the high-quality professional journals.[13]

Many gatekeepers of the higher-quality evaluations stream try to do their part. Out of a concern for external validity, many gatekeepers, including journal editors (Steckler & McLeroy, 2007) and the Cochrane collaborative, have required better and more disclosures of studies' population and sample parameters so that potential users might be able to see how much or how little they resemble the study population. But, while these simple reporting requirements tax the expert evaluators only minimally, they still put a burden on lay users to extract reasonable implications. In most cases, the users cannot do this at all well.

Intermediaries might help. Consider, for example, the International City/County Managers Association (ICMA). They publish a monthly magazine that frequently contains articles of exemplary examples of personnel and sometimes policy from around the country (and sometimes the world). But these are all hopefulness and no skepticism. The worthwhile skepticism of the evaluation experts must therefore be supplied by the readers, but the writing, typically breathless, scarcely encourages them to do so. Face-to-face meetings of local government professionals are likely to host more skepticism, but not, I suspect, enough.

A possibly helpful analogy comes to mind: the "barefoot doctors" in rural settings generally in low-income and low-education countries or regions who function as intermediaries between high-tech medical science and low-tech patient self-care in the rural areas. After some 3–6 months of training, largely in first-aid, emergency, and preventive medicine, these "barefoot doctors" are sent off to "practice," though they clearly lack the sophistication and tools of medical school graduates. However, the success of the barefoot doctors platoons is well documented (Yip, 2018). I will not say much more about this analogy. But think of them as low-tech, but trained and skilled, field-based intermediaries between the ensemble of Elsewheres and the ensemble of Heres, and define their mission in this context.

Reflections on Professional Mission

"Here" and "Elsewhere" are constructs distinctly connected to local government. The "evidence" and "information" in this paper are also much more relevant to local government (and some states) than to national government. Most importantly, the users of the information are assumed to be local officials and their staffs. This is not the sort of clientele assumed by most experts in the evaluation community. This sort of evaluation looks suspiciously like policy analysis (Bardach & Patashnik, 2024).

The policy analysis emphasis is on designing the future rather than critiquing the past. Program evaluation came into its own when, during the Great Society era, even liberals began to question whether the spate of new programs and policies was actually doing much good. Program evaluation shortly came to include a forward-looking component that traveled as "formative" evaluation, in contrast to mere backward-looking "summative" evaluation. If you like, I am recommending an extension of the forward-looking spirit of formative evaluation to an inherently future-oriented, problem-solving, waste-avoiding, opportunity-seizing form of policy analysis.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Eugene Bardach https://orcid.org/0009-0005-0934-0524

Notes

1. A few years after Rossi first asserted his Iron Law, he modified his skepticism somewhat. (Rossi, 2003).
2. To be clear, as a rule, higher-quality information is preferable to lower-quality information. Relying on the results of an RCT is better than relying on a seemingly exemplary example. A locality could run such an RCT, of course, and in many cases probably should, at least eventually. But a big policy and political reason not to do so right away is to avoid losing benefits to the delay that the RCT entails. These benefits would be nil if the tested policy were to prove ineffective; however, they could be substantial if the RCT eventually showed that the intervention was worthwhile. If the jurisdiction in question were set up to run RCT's routinely, and the delays were thereby limited, the risks of running them would be reduced. (Bardach & Patashnik, 2024, p. 86).
3. To be clear, this example is not intended to advocate for or against the Tiny Homes policy. It is intended only to furnish concrete instances of concepts that could otherwise prove too abstract.
4. I have also used (Beasley, 2024).
5. In the Tiny Homes example above, the low-quality before/after performance observations was indeed supplemented by this evidentiary feature.
6. They are also the product of an ontological view that sees the world as made up of zillions of particulars, not just some thousands (?) of general forces. It is more the ontology of biology than of physics or chemistry. These particulars can be thought of as "mechanisms," for which causal understanding must take account of interdependencies among components as well as the forces that help drive them (Glennan, 2017).
7. Implementation-relevant procedures are mainly local but sometimes the state is relevant too, as when the DMV must issue permits to allow trailers to move prefab housing units around. https://richmondconfidential.org/2024/11/25/richmonds-first-tiny-house-project-nears-completion-after-months-of-red-tape/

8. Thomas D. Cook endorses a focus on describing and explaining "causal mediating processes" (Cook, 2014). These seem to me like exactly the kind of processes needed to explicate the workings of "mechanisms."
9. The context problem normally turns up in discussions of that murky concept "external validity." A causal connection shown to exist in one context may not show up in some other context. But it might show up in some others. Hence "external validity" is not an either-or feature of a causal finding, but a matter of degree, with the degree defined by the number and nature of its positive and negative contexts. An exception to this principle occurs when the causal connection exists due to a "natural law" analogous to gravity or Boyle's laws of gases. (Glennan, 2017) The social and behavioral sciences know of no such natural laws, however; at most, energies or forces embedded in "nature," biological or social are the pale simulacra. At any rate, the external validity issue is not relevant for us, as the only external site of concern is "Here," not just a random site plucked from a distribution of possibilities. It is instructive that Bannerjee and Duflo do not refer to the concept in their much-praised book describing anti-poverty RCTs in developing countries (Bannerjee & Duflo, 2011).
10. Often the speaker is right about everything except the "but." Some of Here's differences with Elsewhere will be germane and significant, and others will not be.
11. "Champions" may not be the functional equivalents of "catalysts," however.
12. Well, we hope that many do.
13. An analogous problem affects the hopes of Duflo (Duflo, 2017) for greater enthusiasm among economists to add "plumbing" to their professional brief.

References

Bannerjee, A. V., & Duflo, E. (2011). *Poor economics: A radical rethinking of the way to fight global poverty.* Public Affairs.

Bardach, E., & Patashnik, E. M. (2024). *A practical guide for policy analysis: The eightfold path to more effective problem solving* (7th ed.). CQ Press.

Beasley, S. (2024). Foundations are funding Tiny Home villages for the homeless. *Is it a worthwhile effort?*.

Cook, T. D. (2014). Generalizing causal knowledge in the policy sciences: External validity as a task of both multiattribute representation and multiattribute extrapolation. *Journal of Policy Analysis and Management*, *33*(2), 527–536. https://doi.org/10.1002/pam.21750

DellaVigna, S., Kim, W., & Linos, E. (2024). Bottlenecks for evidence adoption. *Journal of Political Economy*, *132*(8), 2748–2789. https://doi.org/10.1086/729447

Duflo, E. (2017). The economist as plumber. *The American Economic Review*, *107*(5), 1–26. https://doi.org/10.1257/aer.p20171153

Gilbert, D. (2006). *Stumbling on happiness*. Alfred A. Knopf.

Glennan, S. (2017). *The new mechanical philosophy.* Oxford University Press.

Rossi, P. H. (2003). The 'Iron law of evaluation' reconsidered. In Paper presented at the APPAM Research Conference. Washington, D.C.

Schauer, F. (2022). *The proof: Uses of evidence in law, politics, and everything else*. Harvard University Press.

Steckler, A., & McLeroy, K. R. (2007). The importance of external validity. *American Journal of Public Health*, *98*(1), 9–10. https://doi.org/10.2105/ajph.2007.126847

Yip, C.-M. (2018). Off the cuff: 21st-Century barefoot doctors. *Harvard Public Health Review*, *10*.