


ARTICLE

An s-frame agenda for behavioral public policy research

Daniel J. Connolly¹, George Loewenstein¹ and Nick Chater² 

¹Social and Decision Sciences, Carnegie Mellon University, Pittsburgh, PA, USA and ²Behavioral Science Group, Warwick Business School, Coventry, UK

Corresponding author: Daniel J. Connolly; Email: dconnol2@andrew.cmu.edu

(Received 16 October 2024; accepted 22 October 2024)

Abstract

We have previously argued that behavioral scientists have been testing and advocating individualistic (i-frame) solutions to policy problems that have systemic (s-frame) causes and require systemic solutions. Here, we consider the implications of adopting an s-frame approach for research. We argue that an s-frame approach will involve addressing different types of questions, which will, in turn, require a different toolbox of research methods.

Keywords: behavioral economics; behavioral public policy; randomized controlled trials; nudge; behavior change

Introduction

The challenges of public policy are both urgent and diverse: Inadequate retirement income, out-of-control medical costs, rising obesity, political polarization and economic inequality. And, even as these go unaddressed, new challenges constantly emerge. Climate change has already reshaped our habitat and will continue to do so. The COVID-19 pandemic disrupted the social and economic order. Technological innovations, including most recently large language models, are reshaping both work and leisure (Brown *et al.*, 2020; Bubeck *et al.*, 2023). What is the most constructive role that academic researchers, and in particular behavioral scientists, can play in helping society cope with these challenges?

A recent and influential line of work on public policy – arguably the currently dominant viewpoint among behavioral researchers – starts with the premise that adverse policy outcomes frequently result from the frailties of human nature (Camerer *et al.*, 2003; Sunstein and Thaler, 2003). In what Chater and Loewenstein (2023) have termed the ‘i-frame’ (with the ‘i’ standing for ‘individual’), problems such as obesity and insecure retirement are attributed to, for example, *present-bias* – i.e., to people’s tendency to make decisions that yield short-term benefits at the expense of long-run costs.

This perspective has gone hand-in-hand with the notion that public policy can change the ‘choice architecture’ (Thaler *et al.*, 2014) to lead people to make better choices. The solution to insecure retirement, for example, is to encourage people to save more money; the solution to poor public health is to encourage people to improve their diets; and the solution to climate change is to cajole people into reducing their personal carbon dioxide emissions.

Behavioral researchers, who are typically trained to focus on the individual – on how people reason, form beliefs and make decisions – have engaged enthusiastically with the i-frame. In addressing public policy issues such as obesity, retirement security and climate change, researchers have tested a range of diverse interventions to alter the individual behaviors that contribute to these problems.

Given the pervasiveness of what have collectively come to be known as ‘nudges’ (Thaler and Sunstein, 2008), it is not surprising that diverse critiques of this approach have arisen. For example, researchers have raised doubts concerning whether the notion of ‘better’ choices is fully coherent when behavioral factors are accounted for (Infante *et al.*, 2016), whether libertarian paternalism actually meets the definition of paternalism (Mabsout, 2022), whether welfare calculations are possible for policymakers given their information and calculation constraints (e.g., Rizzo and Whitman, 2009, 2019; Sugden, 2018) and about ‘the ethics of nudge’ (Bovens, 2009). In addition, in recent years, two of us (Chater and Loewenstein, 2023) have raised a somewhat different critique: In focusing on helping individuals alleviate their cognitive frailties and behavioral mistakes, the i-frame fails to engage, and indeed distracts from, the ‘s-frame’ (with the ‘s’ standing for ‘system’) – the flawed laws and institutions that are, in fact, largely responsible for almost all the problems that i-frame interventions seek to address.

i-frame policies focus on helping individuals make ‘better’ choices within a given set of rules. These policies largely adopt the approach of altering the presentation of a choice faced by an individual without altering the underlying set of choices: for example, helping workers save for retirement by changing the default option (Beshears *et al.*, 2008), helping consumers choose healthy foods by providing calorie information (VanEpps *et al.*, 2021) or sending criminal defendants reminders to return to court (Fishbane *et al.*, 2020). s-frame public policies, on the other hand, focus on changing the societal rules and institutions which determine the choice sets that individuals ultimately face. Historically, changes in ‘the rules’ have often stemmed from legislation or judicial decisions. These include large transformations, such as socialized medicine, women’s suffrage and gay marriage, but also smaller changes such as environmental fuel standards for vehicles. Technological changes, such as contraceptive pills, the personal automobile and the internet, have also transformed the choice sets faced by individuals. And a third determinant of the choice sets individuals face is ‘social technologies,’ such as labor unions, credit unions and microfinance, but also, more broadly, social and cultural norms, religious beliefs and political partisanship. S-frame behavioral public policy research focuses on these ‘upstream’ domains to understand both the determinants of support for, and the consequences of, different policies.

The i-frame takes the rules and institutions within which people operate as fixed and asks how to help people ‘play’ more successfully within those rules. But when our social and economic games ‘go wrong,’ surely we should be focusing not on fixing the players, but on fixing the rules of the game. From this point of view, a distinctively *behavioral* public policy should focus on how to adapt and design rules that work with the grain of human nature – so that the rules work not merely in principle (e.g., with perfectly rational and informed agents) but when played by people as they actually are, with all of their biases and limitations. This takes us away from individualist solutions, and back to more traditional public policy levers – regulation, taxation, subsidies – although viewed through a behavioral lens.

It is true that the i-frame approach has potentially significant attractions. For example, a commonly cited benefit of i-frame interventions is their cost-effectiveness. As Kahneman (Singal, 2013) expressed it, they can achieve ‘medium-sized gains by nano-sized investments.’ However, even if true (see Tor and Klick, 2022; Thunström, 2019; Thunström *et al.*, 2018 for arguments that the costs of such interventions are higher than has been recognized), the medium sized gains of such policies are typically dwarfed by the scale of the problems they are intended to address. Moreover, recent evidence suggests that the gains from i-frame policies are not, in fact ‘medium,’ but are smaller than most academics believed (DellaVigna and Linos, 2022); and that the effects of i-frame interventions appear larger when they are defined narrowly and on a shorter time frame (Rizzo and Whitman, 2019; Saccardo *et al.*, 2022). Even when they are successful, i-frame interventions are rarely implemented at scale (DellaVigna *et al.*, 2022).

In light of these modest positive impacts, the possibility that a focus on i-frame interventions can ‘crowd out’ more effective s-frame policies looms large. Such crowd-out can occur, first, mechanically, because policy makers and researchers have limited bandwidth – i.e., time, attention and financial resources. Research on, and adoption of, i-frame interventions will inevitably, therefore, come at the expense of competing s-frame policies. Second, crowd-out can occur where the mere availability of i-frame policies reduces support for s-frame policies, both by the public and by politicians, who may see i-frame interventions as cheap alternatives to more costly but vastly more effective s-frame interventions. Indeed, the mere awareness of possible i-frame solutions seems to reduce support for s-frame policies (Werfel, 2017; Hagmann *et al.*, 2019, 2023).

If i-frame policies can serve to crowd out systemic change, then we should expect corporations and wealthy individuals whose bottom-line interests align with the status quo to be among the biggest supporters of i-frame change. And indeed, they are: BP (formerly British Petroleum) coined the ‘carbon footprint’ in 2004 to promote the idea that individual behavior change could address climate change (Solnit, 2021), and the National Rifle Association in America has for decades asserted that ‘guns don’t kill people, people kill people’ (Associated Press, 1987). We have argued that in policy domain after policy domain, these interests have been publicly advancing the i-frame while privately lobbying against s-frame reform (Chater and Loewenstein, 2023). The general principle seems to be that if a group’s aim is to derail systemic reforms, then that group will promote the view that the ‘real’ problem lies with the individual. By focusing

primarily on the i-frame, researchers in the behavioral sciences (ourselves included) have, with the best of intentions, inadvertently helped this process along.

The s-frame is, we believe, where the ‘action’ is. History is a story of continual struggles by different groups within society to control or oppose prevailing rules and institution (e.g., Rizzo, 1980; Acemoglu and Robinson, 2001). Only continual transformation of our systems of rules and institutions has the potential to keep pace with our ever-evolving challenges (North, 1990). Policy debates must be cast in the context of political economy: We need to ask which interests are impacted by engaging in (or blocking) reform, and how those interests are attempting to control the political process.

While these concerns over i-frame policies have received considerable recent attention, our focus here is not on the policies themselves, but on the implications of adopting an s-frame perspective for behavioral public policy *research*. Shifting our focus toward the s-frame – to the goal of figuring out how to create beneficial systemic change – will, we argue, require the behavioral research community to address a new set of questions, which will, in turn, require a different mix of research methods. Our aim is, therefore, to provide an agenda for behavioral s-frame research. Improving policy design, implementation and communication will entail trying to understand how behavioral agents will respond to changes in the underlying systems in which they operate. We discuss, first, the new types of research questions to be addressed, and, second, the different mix of research methods that will be required.

Asking new questions

The transition from i-frame toward s-frame research will, most prominently, involve a change in the mix of questions addressed by researchers. In this section, we discuss key behavioral public policy research questions, engagement with which could advance an s-frame research agenda: For a given issue area, how to improve policy *design*; *implementation*; and *communication*, and, at a broader level, how to consider the behavioral factors that generate large-scale social change.

Better policy design

As we’ve noted, i-frame policies typically focus on ‘choice architecture,’ interpreted as applying to the end-line decisions faced by consumers, taking the wider system as given. For example, an i-frame approach to diet might involve changing the arrangement of food in a cafeteria or labelling food with calorie counts. But modifications of choice architecture can be applied to aspects of the choice environment which are more deeply embedded in the system itself: The taxes and subsidies which govern the relative prices of food items in a store, regulations on the formulation of those items (e.g., quantities of salt and sugar), or the zoning and planning decisions which affect where that store is located. These topics are of intense interest to economists and sociologists but have been largely neglected by the behavioral public policy community. There is no reason, however, why behavioral scientists shouldn’t play a central role in the ‘whole-cloth’ design of s-frame policies to maximize their effectiveness, increase their appeal and defend them from political attack.

Indeed, one of behavioral public policy's earliest successes, opt-out 401(k) contributions in the United States, entailed a change in the legal framework surrounding American retirement accounts, leveraging the behavioral insight that American workers were unlikely to deviate away from a default. Yet insecure retirement remains a problem in the United States for many workers. There is ample opportunity for a more expansive analysis of how the United States might find its way to comfortable retirement for all, a state of affairs already achieved by many other countries. Such fundamental change is needed, even according to some of the central contributors to the original 401(k) default contribution research (Laibson, 2020; Choi *et al.*, 2023).

A behavioral approach to the problem of financial insecurity in retirement would identify policies that could address the shortfall in retirement provision: Policies that are comprehensible to individuals (who are both voters and current or prospective retirees), and that are practical – i.e., potentially implementable by the U.S. government. Many policy ideas have been proposed: The 'auto-IRA' (universal automatic enrollment into a payroll-deduction IRA; Iwry and John, 2021), expanding Social Security benefits (Goss, 2023) and mandatory contributions (Beshears *et al.*, 2020). The ideal s-frame research program for secure retirements would define a normative criterion for 'solving' the problem of funding retirement, or at least set a minimum viable goal. Defining this criterion is the first area in which the behavioral sciences are equipped to help develop policy. Supposing that the preferences of voters should help define a goal for optimal retirement policy, one challenge in defining this goal is that the preferences of voters over retirement policies may change *ex post* policy reform. A further problem is that, if preferences depend on the choices people are offered (Bernheim *et al.*, 2024), the policy preferences elicited from voters may vary with the options that are 'on the table.' Rather than seeing these issues as insurmountable, we view them as an exciting challenge to the behavioral public policy community. It may even be the case that the behavioral public policy community embraces the need to persuade the public about the ideal secure retirement policy; this would require that behavioral public policy researchers select their preferred criterion and justify it.

Once a normative criterion has been defined, a set of existing and new ideas could be examined for their (i) potential to meet this goal, taking account of individual psychology (i.e., how people will react to the rules, including changing other aspects of their spending and saving) and their (ii) potential to be implemented by political institutions (in this case, in the United States). While we do not take a position on what specific ideas should be examined, behavioral science suggests some obvious features. For example, if the system involves the accumulation of 'pension pots' (as in the UK), people should not be asked to make decisions about how to invest their own funds, given the low level of financial literacy among the U.S. population and the virtual impossibility of educating people to the point of being able to make such decisions in an informed fashion (Willis, 2008). Behavioral research into alternative retirement systems could include stylized survey-based experiments to examine relative support for different policies, comparative (including quasi-experimental) research on policy changes in other nations and theoretical models and agent-based simulations of the general-equilibrium economic and political effects of changes in U.S. retirement policy.

In some cases, the goals of maximizing a policy's public appeal and optimizing its impact will be in tension. For example, a consumer-facing carbon tax – administered

on goods and services based on the carbon gas emissions they entail – would likely be most effective if separated from the cost of the good or service itself. Shoppers might be deterred from purchasing highly taxed goods by the thought that a chunk of their money covers the tax and is not indicative of the value of the good itself. But a more salient tax might thereby also be a less popular tax (Bracco *et al.*, 2019). People seem to prefer ‘hidden’ taxes – like the VAT in Europe – for which it is unclear who are shouldering the burden (of course, these conjectures themselves are good targets for behavioral testing.) In other situations, however, the goals of acceptability and impact may be aligned. For example, people seem to prefer, perhaps surprisingly, that the revenue from a carbon tax be used to subsidize climate-change-combating activities (Amdur *et al.*, 2014; Reed *et al.*, 2019; see also Marron and Morris, 2016). While specific examples exist of policies that have been both widely accepted and efficacious, we lack a formal framework that could provide systematic guidance on how to design a policy to maximize support without compromising its intended aims.

Better policy implementation

Even if a policy is designed to function optimally for behavioral agents in equilibrium, most policies will require careful introduction and continual adjustment because equilibrium behavior under a policy may differ from initial behavior. For example, as targets of a new policy learn how it works, they may become more sophisticated about its purpose, and, as a result, become more (or less) supportive. But the nature of this process for any particular policy is not obvious. As Rizzo and Whitman (2018, p. 215) note, ‘In new and unfamiliar contexts, especially, people will make systematic errors for a certain period of time. The precise length of time it takes to learn will vary from problem to problem and environment to environment as well as with the learning propensities of the individual.’

Two behavioral research topics with especially important implications for how people will respond to changes in policy are *adaptation* and *learning*. Research on adaptation shows not only that people adapt rapidly to an astoundingly wide range of conditions (including adverse outcomes as bad as paraplegia; Brickman *et al.*, 1978; Frederick and Loewenstein, 1999) but also that people underpredict their own speed of adaptation (Ubel *et al.*, 2005; Loewenstein and Ubel, 2008). Research on learning likewise finds that people can learn to operate with relative efficiency in an extraordinarily wide range of environments, and that they tend to learn much more rapidly than they anticipate (Koriat *et al.*, 2002; Billeter *et al.*, 2011).

Many of the policies required to address social challenges are new, so past experiences in one’s own or other societies will be of limited value. Nonetheless, the research findings on adaptation and learning point to an important regularity we should expect to observe: Learning how to deal with, and adapting hedonically to, new policies, should be more rapid than people anticipate at the outset. Consider, for example, shifts to compulsory seatbelts, bans on smoking in public places or radically reducing alcohol limits for drivers. These, and many other, regulatory changes all met initially with vigorous opposition from relevant lobby groups, the press and the public. But people seemed to rapidly adapt to these restrictions, so that what might initially have been

perceived as an infringement of liberty was soon mostly viewed as a mutually agreed constraint for the common good (Diepeveen *et al.*, 2013; Phanér and Hane, 1979).

A clever and especially compelling illustration of this pattern comes from a laboratory market experiment by Janusch *et al.* (2021; see, also, Cherry *et al.*, 2012) in which participants initially opposed a stylized congestion pricing policy but came to support it once they had experienced its efficiency-enhancing effects. An important implication of this pattern for research is that initial measures of people's preferences may give a misleading picture for policymakers about the long-term acceptance of policies. This observation raises a range of practical questions for policy introduction: Under what conditions should policies be introduced abruptly vs gradually? How far in advance should changes be announced? How light-touch or heavy-handed should enforcement be (to encourage compliance, but avoid public rejection)? Mistakes in policy implementation can be costly – witness the political backlash to taxation on gasoline illustrated by France's 'yellow vest' movement, and controversy generated by plans for so-called Ultra-Low Emissions Zones in UK cities (Wills, 2023). The forces determining the difference between public acceptance and rejection are largely behavioral. Thus, behavioral researchers can and should develop a scientific understanding of these forces.

Better policy communication

Often there is a near consensus about which s-frame solutions will successfully address a problem (e.g., because these solutions have proven successful where they have been adopted). There are, for example, a range of different and broadly successful approaches to retirement funding, and to the universal provision of health care. But despite their proven success, these policies are often not adopted when and where they are needed. Improving communication about such s-frame policies – including explanations of how and why they work, as well as documentation of their success – is an area in which behavioral public policy researchers can play a pivotal role. Unlike behavioral policy design and implementation, understanding how to increase understanding of, and support for, proven policies is a research topic for which the current suite of experimental tools from behavioral science (laboratory experiments, randomized controlled trials [RCTs]) are, in fact, well-suited (and can complement typically more qualitative analysis in neighbouring disciplines, such as the communication, political and language sciences, e.g., Lakoff, 2010; Pezzullo and Cox, 2022).

Building support for needed policies means, in many cases, helping people to understand how and why those policies work. People will not support beneficial policies that they can't understand and, worse, they are likely to support ineffective policies that they incorrectly believe they do understand. A paradigmatic recent example of the latter was the Brexit campaign, which the right-wing media sold with a combination of emotion-evoking untruths about the supposed negative impact of the EU (e.g., on immigration, the national health service and autonomy) and rosy projections of the economic benefits of exiting the EU (Atikcan *et al.*, 2020). The disastrous U.S. response to the COVID-19 pandemic provides another illustration: Officials at the Centers for Disease Control and Prevention recommended against wearing facial coverings in March 2020 but reversed their position in April. By July, White House

Coronavirus Task Force member Anthony Fauci was quoted as saying, ‘We have to admit it, that that mixed message in the beginning, even though it was well meant to allow masks to be available for health workers, that was detrimental in getting the message across’ (Breslow, 2020). Examining how to best communicate the rationale behind policy decisions may help enhance the effectiveness of those decisions. Thus, if the public had understood that recommendations against masking in March 2020 stemmed from concerns about the supply of masks and not from concerns about their effectiveness, masking may have been more widespread after the CDC reversed its recommendation.

The behavioral science tools for building policy support via better communication are largely in place. We know, for example, that people make sense of many aspects of the world in terms of *narratives* (e.g., Johnson *et al.*, 2023), and a familiar challenge is guiding people to adopt a narrative favorable to a product or service being marketed (Woodside *et al.*, 2008; Dessart and Pitardi, 2019). When communicating about policies, understanding the behavioral factors that shape policy framing is therefore of crucial importance. This is familiar territory for the study of party-political campaigning and messaging concerning policies, from individual policies to entire political platforms. Psychological principles are clearly central to such an understanding. For example, attentional limitations may force us to use simple heuristics (such as attribute substitution [Kahneman and Frederick, 2002], where a difficult question is replaced by a simpler one); or might drive us to rely on social proof (following the thoughts and behaviors of trusted leaders or peers [Cialdini and Goldstein, 2004]). Moreover, people may be more concerned with working out what their peer group will view as ‘appropriate’ (i.e., do ‘we’ collectively acquiesce or rebel to a policy change), rather than evaluating the consequences of a policy (March and Olsen, 2008). Note that applying these lessons to substantively increase public support for effective policies will require overcoming opposing messages coming from entrenched economic and political interests, forces that the behavioral public policy community has yet to confront effectively.

A cynical interpretation of ‘better policy communication’ is, of course, possible: That these efforts are behavioral and cognitive ‘tricks’ intended to manipulate public opinion. Behavioral public relations or propaganda campaigns do occur and are not sincere attempts to communicate; indeed, they often covertly aim to obscure and mislead, such attempts at misleading the public should be studied primarily so that they can be opposed. The goal of genuinely better policy communication is very different: To enhance public comprehension of policies that are empirically supported and likely to be effective. To the extent that there is disagreement between the policy views of researchers and the public, we believe that researchers should be willing to defend the normative criteria on which their own policy views rest. Indeed, they should be willing to attempt to persuade the public to share their views – to play a part in the public debate.

The bottom-up forces that drive social change

Rizzo and Whitman (2019; see also Sugden, 2018) worry that work in behavioral public policy frequently seems to presuppose that the objectives of policy are set

from 'outside' mainstream society (for example, by policymakers who presume to know the public's interests better than the public does). This top-down perspective fails to capture the fact that profound social changes, such as the end of slavery, women's suffrage, the right of workers to unionize, the civil rights movement and legislation outlawing discrimination based on gender and sexuality, are often brought about by long periods of grassroots campaigning by oppressed groups and their allies.

The question of when and how bottom-up social movements create change has, of course, been a major focus of research in political science, sociology and history (e.g., Tilly, 2004). But many specifically behavioral questions arise: How do changing attitudes and beliefs propagate across communities? How does the pressure of social conformity squelch the expression of new ideas and attitudes, and how can such pressure be overcome? What are the psychological and organizational factors that drive people to campaign and protest in the face of a difficult collective action problem, in which the costs of campaigning are borne by individuals in the short-term, but any benefits they accrue are diffused across society and across time?

Understanding society's capacity for bottom-up change will require taking account of important elements of individual psychology, many of which may emerge from our evolutionary history in relatively stable small groups, rather than large and complex societies (Bowles and Gintis, 2011). People's moral values, including those based on allegiance, loyalty, purity, fairness, equality and justice, in addition to their concerns about outcomes (Haidt, 2008), the high value that people place on others sharing their beliefs (Golman *et al.*, 2016), and the formation of individual and collective identities (Klandermans, 2014) may all be crucial to understanding the origin and spread of social movements. Another important ingredient in the emergence of social phenomena from individual psychology is people's lay narratives about economy and society (Andre *et al.*, 2022; Johnson *et al.*, 2023), and how these narratives are created or subverted top-down by lobbyists, think tanks and media providers aiming to promote the interests of wealthy and powerful individuals and corporations (Walker, 2014; Oreskes and Conway, 2023).

To date, the behavioral public policy community has largely focused on evaluating the potential for individual interventions to change individual behavior. We believe that researchers should attempt to maximize public support for effective policies via the tools of policy design, implementation and communication. We also believe that researchers should attempt to understand how dramatic policy shifts can and do emerge from mass mobilization absent a technocratic policy development process. Asking these new questions would constitute a needed shift in the focus of the behavioral public policy community.

Broadening the range of research methodologies

Asking different questions will necessarily entail employing different research methods. Here, we advocate for a shift from field experimentation (which are often ideal for testing i-frame interventions) toward quasi-experimental and even qualitative observational approaches examining changes in policies across countries and over time.

Dethroning the RCT

Currently, there appears to be a virtual consensus among practitioners of behavioral public policy that RCTs are the best methodology for testing new policy interventions (Haynes *et al.*, 2012; see Luca and Bazerman, 2021, for an entire book making such a case, and Duflo and Kremer, (2005) for a similar perspective from development economics). Indeed, when advocates of i-frame approaches have come to recognize the limited impact of the interventions they have historically supported, they sometimes seem to retreat from advocating specific interventions to the ostensibly more impregnable position that the essence of the behavioral approach is that policies should be empirically tested with RCTs. Here, we challenge this seemingly uncontroversial perspective.

A number of criticisms have been leveled against enshrining RCTs as the gold standard (Frieden, 2017). Deaton and Cartwright (2018), for example, focus on two broad limitations, which – though they studiously avoid the terms – roughly correspond to internal validity (the ability to draw confident conclusions about causal effects) and external validity (the ability to generalize/extrapolate from findings to broader contexts of interest). Our critique is different, and more in line with a later influential paper by Deaton (2020) provocatively titled ‘Randomization in the Tropics.’ In a subsection of this paper titled ‘Small versus Large,’ Deaton argues that most of the changes in economic policy that have led to substantial improvements in economic well-being have resulted from ‘large’ shifts in policy, but that RCTs are limited to testing the impact of ‘small’ innovations (see, also, Rodrik, 2009). RCT tests of policy interventions are constrained to those that can be randomized to a unit of analysis, most commonly the individual (Bédécarrats *et al.*, 2019). Green energy nudges can be tested with an RCT; a carbon tax cannot. Interventions to help people make better health insurance choices can be randomized; single payer insurance cannot. Modified litter bins (e.g., with ‘watching eyes’) can be trialed using RCTs (Bateson *et al.*, 2015); sweeping and complex regulations based on, for example, Extended Producer Responsibility to reduce packaging cannot (Coombe, 2023).

The requirement that policy interventions should be validated with an RCT serves, therefore, to drastically limit the range of policies that can even be considered. Even when it is conceivable that an s-frame policy *could* be tested with an RCT, practical considerations almost invariably prevent that from happening. As Thaler (2023) notes (interestingly, in a critique he wrote of Chater and Loewenstein’s, 2023 i-frame/s-frame paper),

the range of interventions studied by behavioral scientists is truncated by what I call permission bias: You can only test what you can get the approval to try. It is wrong to infer from this fact of life that behavioral scientists are using the wrong “frame”. Rather, they face constraints! It also makes it problematic to judge the potential impact of possible behavioral policy interventions based on the set of randomized control experiments behavioral scientists have been allowed to run.

It is, indeed, a ‘fact of life’ that conservatism on the part of policy makers, as well as cost considerations, will limit the types of policies that get tested to those that are cheap and easy to implement, such as changes in emails or mailings. Thus, if testing

with an RCT is a requirement for implementing a policy, then the ‘permission bias’ that Thaler mentions will restrict policy options to those that can be *tested* cheaply and easily. As Jackson (2023) writes, ‘It is hard to imagine how the welfare state reforms of the last century could have been introduced under current evidentiary standards for policy implementation: These reforms were simply too expansive in their scope, and there was certainly no body of experimental or quasi-experimental evidence to support an overhaul of multiple institutions.’ Deaton likewise notes that a wide range of policy reforms that led to dramatic improvements in quality of life, such as the introduction of markets in China, which lifted a large fraction of the world’s population out of poverty, could not have occurred if they first had to be tested with RCTs prior to implementation.

While running experiments, much like i-frame policies themselves, might seem uncontroversial, to the extent that experimentation is viewed as the gold-standard, there will be a tendency not only to devote resources toward RCTs that could go toward other methods, but, perhaps even more seriously, to downplay the validity or relevance of results obtained using other methods. Enshrinement of the RCT will, therefore, inevitably direct attention away from policy interventions for which RCT evidence cannot be obtained. In this way, the elevation of experimentation above other methods can raise profound ethical issues. Deaton (p. 45) echoes Thaler’s complaint of ‘permission bias,’ but draws almost the opposite conclusion,

In authoritarian regimes with full control, it is only possible for outsiders to help when it is in the government’s interest to accept that help. Development agencies then find themselves in the situation of being “allowed” to help the poor, or to help provide health services, while providing political cover for the “enlightened” despot who is thereby free to persecute or eliminate his opponents. Similar issues arise in democracies too, though less sharply; the step from evidence to policy is never ethically neutral but is less fraught when the poor have a voice and some political power.

This is not to argue that RCTs should play no role in s-frame-oriented policy research. For example (as discussed above), in cases where strong policy solutions exist but are constrained by public support, experimentation on how to maximize public support by re-framing or redesigning policy elements is likely to be highly productive. But experiments alone are unlikely to be able to answer many other important policy questions such as: Which of two potential policy solutions is likely to lead to the largest increase in welfare? Which factors best predict whether an economically efficient policy will be politically stable? Why are good policies adopted in some countries and not others?

Embracing a greater diversity of methods, including methods that are already widespread in economics and sociology but do not offer the same degree of causal identification as RCTs (such as event studies, regression discontinuity and synthetic control studies), as well as methods from other fields that used to have more of a home in the behavioral sciences (such as historical and comparative research), will enable behavioral public policy researchers to engage more fully and productively with system-level policy research.

Quasi-experimental methods

As RCTs have become popular in economics, so too have economists made advances in *quasi-experimental methods* for causal inference. In limited cases, in fact, something quite close to experimentation occurs ‘in the wild.’ Most obviously, sometimes there is actual random assignment done for purposes other than research, such as fairness – e.g., the draft lottery (Angrist, 1990; Angrist and Krueger, 1995) and random assignment of cases to judges (Green and Winik, 2010; Dobbie *et al.*, 2018). In other cases, policy, or other, changes occur at some point in time in one sample but not in other, roughly comparable ones, enabling a difference-in-difference approach (e.g., Wing *et al.*, 2024). In still other cases, there is a sharp cutoff, e.g., by age, or test score, in who receives different treatments, enabling regression discontinuity analyses. Such quasi-experiments are not methodologically superior to RCTs, but they have different limitations and scope, and hence should be embraced by behavioral public policy researchers. Perhaps the main difference between experiments and quasi-experiments is that the latter are often well-suited to analyzing the impact of real-world problems and the policies designed to address them. For example, these methods have been used to analyze both the effects of air pollution (Currie *et al.*, 2015) and the effects of a cap-and-trade market designed to limit air pollution (Deschênes *et al.*, 2017).

Researchers have ample opportunity to adopt quasi-experimental methods and turn them towards behavioral public policy questions in the s-frame: How has public opinion and policy support for action on climate change changed in the wake of severe weather events? How does support for congestion pricing change after bad weather or accidents cause particularly bad traffic? Questions like these are already of interest to behavioral public policy scholars, are useful for the s-frame agenda and can be answered using quasi-experimental methods.

Historical and comparative methods

While some policy challenges are new, many are not, and no challenge is entirely novel. *Historical methods* are robustly employed in the subfield of economic history; economists and historians employ economic models and archival research methods to generate and test hypotheses about how economic activity generates and is generated by historical events. By analogy, ‘psychological history’ can examine how historical forces shape individual psychology (e.g., beliefs, attitudes, preferences, norms) and how changes in individual-level psychological constructs can influence historical events. We are not the first to make this point (see Seligman, 2023; Baumard, 2019; Murthukrishna *et al.*, 2021), but we believe that ‘psychological history’ is a relatively untapped direction for generating and testing hypotheses about the structural policy questions that face us today. For example, Hargreaves Heap (2024), postulates that two major real-world macroeconomic phenomena (declines in productivity growth and increases in wage inequality in rich countries over the last 50 years) can be traced to changes in the risk preferences of individuals in those countries. Linking individual psychological constructs with broader structural change in this way could help researchers to answer new questions about how novel policies come to be enacted.

The need for historical approaches also coincides with novel machine learning techniques for investigating the psychological and social constructs within significant

historical periods through automated text analysis. These techniques have enabled researchers to use archival text to understand peer influence during the French Revolution (Barron *et al.*, 2018), the concept of ‘rationality’ during the Industrial Revolution and its aftermath (Scheffer *et al.*, 2021) and the women-led development of abolitionist arguments in pre-emancipation America (Soni *et al.*, 2021). Corpora of text in marketing and political campaigns could similarly help policy researchers understand the development of successful policy initiatives like smoking bans and seatbelt laws.

Just as we can look to history, we can look across the world to understand how different polities have achieved different policy outcomes. A particularly useful econometric technique is the creation of a ‘synthetic control group’ from a weighted sum of comparator units, which can then be compared against the treatment unit on an outcome of interest. Synthetic control can be employed to compare the causal effect of various shocks across cities (Peri and Yasenov, 2019), regions (Abadie and Gardeazabal, 2003) and nations (Billmeier and Nannicini, 2013). Beyond specific techniques, researchers should use both qualitative and quantitative methods of comparison to address s-frame questions: How do different political, economic and media institutions generate different beliefs among the public about policy topics? How might changes in these institutions effect changes in beliefs?

The need for learning in behavioral policy development

In addition to embracing a broader variety of existing methods in the behavioral sciences, we believe that researchers and policy makers should more explicitly embrace a ‘learning agenda’ of policy development. The psychology of human learning and the parallel discipline of machine learning provide a powerful counterweight to the current emphasis on RCTs and other controlled experimentation and permit an observational approach that extends beyond a single historical or cross-national comparison. RCTs are designed to test which of two or three fully fledged variations is most successful (e.g., drug vs placebo; tax letter X, Y or Z) – a direct test that quasi-experimental, historical and comparative methods can only approximate. But s-frame policy efforts require creating and continually adjusting complex and highly interdependent programs of actions and messages; these programs operate on an unstable and reacting world which cannot be navigated with a tree of binary or ternary choices. We are not the first to make this point: Newell (1973) makes a parallel critique of experimental methods in cognitive psychology, though with a rather different focus, Irzik (1985) makes a related critique of Popper’s advocacy for ‘piecemeal engineering’ (1957), and Hausmann and Rodrik (2003) make a similar point in the context of economic development.

Such a ‘trial-and-error’ approach accords with how people learn to maximize their own welfare. In a discussion of how to evaluate the welfare maximization of individuals, Rizzo and Whitman (2018, pp. 216–7) point out that,

Real human beings do not make decisions instantaneously and without error. Nor do they know all of their goals, and their fully specified willingness to trade them off against each other, prior to making any decisions. Nor do they hold beliefs that are instantaneously consistent with each other and the world.

Forming one's preferences and beliefs is a process, and therefore it seems natural to evaluate them in terms of that process.

As researchers, and policymakers, we should expect that to understand and interact in real time with a complex and changing world, we will need to use both unsupervised learning strategies (seeking patterns in observational data) and reinforcement learning strategies (dynamically iterating toward a defined goal), rather than being restricted to a narrow subset of 'methodologically pure' methods which capture a static understanding of the subjects of a policy.

Indeed, the vast bulk of human learning is what is known as 'unsupervised' learning, in which an intelligent agent finds patterns in sensory or linguistic data. One approach, for example, is to build a model that successively predicts new data from prior data, as in the deep neural networks that have recently been revolutionizing artificial intelligence (see, e.g., Brown, Mann, Ryder, Subbiah, Kaplan, Dhariwal, Neelakantan, Shyam, Sastry, Askell *et al.*, 2020, Vong *et al.*, 2024). Another related approach is to attempt to *compress* linguistic or sensory data through a bottleneck in a neural network (e.g., Hinton and Salakhutdinov, 2006), or to find symbolic representations that capture the structure in data in a highly compact form (e.g., Griffiths *et al.*, 2010; Yang and Piantadosi, 2022). Yet another approach is to use causal Bayesian networks to find causal structure from observational, not experimental, data (e.g., Gopnik *et al.*, 2004; Pearl, 2009; Chater and Oaksford, 2013). In short, the fields of cognitive science and artificial intelligence suggest that while controlled experiments may be useful where they can be carried out, almost all human intelligence and ingenuity operates using entirely different methods. Thus, a perhaps unexpected lesson for behavioral public policy is that the success of human intelligence in dealing with an immensely complex natural and social world is itself a powerful demonstration that a restriction to the gold standard of RCTs and laboratory experimentation is both drastically over-restrictive and unnecessary.

A further lesson from the study of human and machine learning is that our best theories of how intelligence systems can learn to *control* a complex world operate not by ever more refined RCTs, but through dynamic interactions with that world, and successively updating the policies governing those interactions using principles of reinforcement learning (Rescorla and Wagner, 1972; Hausmann and Rodrik, 2003; Sutton and Barto, 2018): Essentially a process of continuous 'tinkering' while exploring a 'policy' (which, here, may be a set of strategies governing a person's actions), through trial-and-error. Rizzo (2021) discusses a similar idea for the development of social systems of justice: 'Rules [of justice] do not come into being immediately ... they come into existence as the result of a trial-and-error process. Individuals learn that to more effectively pursue their self-interest they must restrain self-interest by adhering to certain rules.'

Reinforcement learning identifies patterns of interaction with the environment that lead to positive or negative outcomes and successively modifies the agent's policy to improve the average level of reinforcement received. No RCTs are used – even though the problem is directly to learn to *control* the environment. Yet such methods, and especially so-called 'deep' reinforcement learning in which underlying representations of the environment are not fixed, but are learned through neural network learning,

have been spectacularly successful in learning to play at, or beyond, human levels of performance in domains as varied as Atari video games, chess and Go (Schrittwieser *et al.*, 2020); and related methods have even been applied in crucial breakthroughs in predicting protein folding from DNA sequences (see Callaway, 2020). Behavioral public policy research can draw on sophisticated reinforcement learning methods, which are beginning to be adopted in econometrics (Athey and Imbens, 2019).

The broader insight to draw is that practical policymakers need not ‘turn off’ their natural modes of intelligent engagement with the world and rely purely on RCTs. In the real-world of business and government, policymakers continually draw on intuitions and insights from beyond the restrictions of knowledge gained from RCTs – they employ the full range of strategies that make human intelligence successful in dealing with an often overwhelmingly complex world. Of course, intuition can mislead; critical analysis and carefully curated evidence is vital; and experimentation can play a key role in teasing out causal mechanisms. But if researchers and policymakers adhere rigidly to the gold standard of RCTs, they will be foregoing almost all the techniques that make human intelligence successful.

Conclusions

The daunting challenges facing society can seem unprecedented and insurmountable. But history tells a different story: Humanity has faced a continual stream of challenges, from devastating famines, droughts, endemic malnutrition, vulnerability to disease, global wars, invasions, political oppression, revolutions, civil unrest and even past climate change (Parker, 2013; Fagan, 2019). While many of the challenges remain, huge progress has been possible – one index being the remarkable increase in human life expectancy (Roser *et al.*, 2013). Such progress has not primarily been generated through direct shifts in individual behavior – rather, it has been created by radical systemic reforms to political, economic and legal systems and institutions, alongside dramatic technological and scientific advances (e.g., North and Thomas, 1973). But institutional change is not guaranteed to lead to progress; it sometimes leads to social and economic decline (Acemoglu *et al.*, 2002). Which reforms ‘win through’ depends in large part in the proper functioning of the marketplace for ideas, and the processes of debate and deliberation which can turn those ideas into reality. A well-functioning marketplace will depend on the quality of its regulation and its robustness to subversion by powerful special interests who gain from the status quo, and which have historically resisted, and will doubtless continue to resist, change for the common good (Acemoglu *et al.*, 2013). The marketplace for ideas cannot be left to the market – behavioral public policy research can make a crucial contribution to understanding what needs to be done.

Given the focus of this article, it would be a mistake to conclude only by exhorting individual researchers to pursue the s-frame agenda. Shifting behavioral public policy research toward s-frame applications also requires a change in research culture and norms – moving beyond tightly controlled laboratory experiments and in-field RCTs to the full range of strategies by which human intelligence grapples with a complex and uncertain world. Changing culture in this way will require several systemic changes in the research culture itself. Broadening the methodological training

of graduate students will require coursework that goes beyond experimental methods to include quasi-experimental, historical and comparative methods. Journals should develop guidance for how to successfully publish s-frame research: Standardized methods for comparing not just effect sizes but scope for total impact, a tolerance for trading off strict assumptions of causal identification for an accumulation of less-causal evidence, and special issues that focus strictly on s-frame research. And policymakers and professional organizations should support the s-frame agenda, financially and publicly, by integrating researchers into their operations and supporting administrative data access for s-frame research. While academic research is only of many forces shaping political debate, these reforms will help behavioral public policy researchers contend with the corporate interests that have successfully framed public policy debates in individualistic terms.

Beyond these institutional changes, there needs, too, to be a shift toward encouraging and rewarding imagining radical possible policies and futures (Mulgan, 2022). The invention or modification of new rules and institutions is surely at least as important for the social sciences as the creation of conventional technology (from engineering, computer science or pharmaceuticals) has been for the natural sciences, and yet social science has generally failed to engage the challenge of ‘social technology’ in the same fashion. A focus on conceiving and evaluating radically different policy proposals will pull the field away from theoretically ‘cute’ findings and the search for rigorously evidenced ‘tweaks.’ Instead, we believe that there should be an imperative to focus on understanding systemic changes, and the processes by which those changes are framed and debated, and that this imperative will be required to create a future that promotes, rather than undermines, the common good.

Acknowledgements. We thank Erin Sherman for helpful comments on a previous version of this manuscript. NC gratefully acknowledges support from the European Union/UKRI under Horizon Europe Programme Grant Agreement no. 101120763 – TANGO. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Health and Digital Executive Agency (HaDEA). Neither the European Union nor the granting authority can be held responsible for them. For the purpose of open access, the author has applied a Creative Commons Attribution (CC-BY) license to any Author Accepted Manuscript version arising from this submission.

References

- Abadie A. and J. Gardeazabal (2003), ‘The economic costs of conflict: a case study of the Basque country’, *American Economic Review*, **93**(1): 113–132.
- Acemoglu D., S. Johnson and J. A. Robinson (2002), ‘Reversal of fortune: geography and institutions in the making of the modern world income distribution’, *The Quarterly Journal of Economics*, **117**(4): 1231–1294.
- Acemoglu D., S. Johnson, and J. A. Robinson (2013), *Why Nations Fail: The Origins of Power, Prosperity, and Poverty*, New York, NY: Crown Currency.
- Amdur D., B. Rabe and C. P. Borick (2014), ‘Public views on a carbon tax depend on the proposed use of revenue’, *Issues in Energy and Environmental Policy*, **13**.
- Andre P., C. Pizzinelli, R. Roth and J. Wohlfart (2022), ‘Subjective models of the macroeconomy: evidence from experts and representative samples’, *The Review of Economic Studies*, **89**(6): 2958–2991.
- Angrist J. D. (1990), ‘Lifetime earnings and the Vietnam era draft lottery: evidence from social security administrative records’, *The American Economic Review* **80**: 313–336.
- Angrist J. D. and A. B. Krueger (1995), ‘Split-sample instrumental variables estimates of the return to schooling’, *Journal of Business & Economic Statistics*, **13**(2): 225–235.

- Associated Press (1987, September 22), 'Britain to tighten gun laws after Hungerford massacre'. *AP News*. <https://web.archive.org/web/20220707130007/https://apnews.com/article/d23e5062a060cfb495662c5b21cf3642>
- Athey S. and G. W. Imbens (2019), 'Machine learning methods that economists should know about', *Annual Review of Economics*, **11**: 685–725.
- Atikcan E. Ö., R. Nadeau and É. Bélanger (2020, November 3), 'Framing Risky Choices: How the Leave Campaign Convinced Britain to Take a Leap into the Unknown'. *EUROPP –European Politics and Policy*. <https://blogs.lse.ac.uk/europpblog/2020/11/03/framing-risky-choices-how-the-leave-campaign-convinced-britain-to-take-a-leap-into-the-unknown/>
- Barron A. T. J., J. Huang, R. L. Spang and S. DeDeo (2018), 'Individuals, institutions, and innovation in the debates of the French Revolution', *Proceedings of the National Academy of Sciences*, **115**(18): 4607–4612.
- Bateson M., R. Robinson, T. Abayomi-Cole, J. Greenlees, A. O'Connor and D. Nettle (2015), 'Watching eyes on potential litter can reduce littering: evidence from two field experiments', *PeerJ*, **3**: e1443.
- Baumard N. (2019), 'Psychological origins of the industrial revolution', *Behavioral and Brain Sciences*, **42**: e189.
- Bédécarrats F., I. Guérin and F. Roubaud (2019), 'All that glitters is not gold. The political economy of randomized evaluations in development', *Development and Change*, **50**(3): 735–762.
- Bernheim B. D., K. Kim and D. Taubinsky (March 2024), 'Welfare and the Act of Choosing', NBER Working Paper No. w32200, Available at SSRN: <https://ssrn.com/abstract=4746810>
- Beshears J., J. Choi, C. Clayton, C. Harris, D. Laibson and B. Madrian (2020), Optimal Illiquidity (W27459), National Bureau of Economic Research.
- Beshears J., J. J. Choi, D. Laibson, and B. C. Madrian (2008), 'The importance of default options for retirement saving outcomes: evidence from the USA', in Kay, and Sinha (eds), *Lessons from Pension Reform in the Americas*, New York: Oxford University Press, 59–87.
- Billeter D., A. Kalra and G. Loewenstein (2011), 'Underpredicting learning after initial experience with a product', *Journal of Consumer Research*, **37**(5): 723–736.
- Billmeier A. and T. Nannicini (2013), 'Assessing economic liberalization episodes: a synthetic control approach', *The Review of Economics and Statistics*, **95**(3): 983–1001.
- Bovens L. (2009), 'The ethics of nudge', in *Preference Change: Approaches from Philosophy, Economics and Psychology* Grüne-Yanoff Till, and Hansson Sven Ove eds, Dordrecht: Springer Netherlands, 207–219.
- Bowles S. and H. Gintis (2011), *A Cooperative Species: Human Reciprocity and Its Evolution*, Princeton, NJ: Princeton University Press.
- Bracco E., F. Porcelli and M. Redoano (2019), 'Political competition, tax salience and accountability. Theory and evidence from Italy', *European Journal of Political Economy*, **58**: 138–163.
- Breslow J. (2020), 'Fauci: Mixed Messaging on Masks set U.S.' Public Health Response Back. *National Public Radio*. <https://www.npr.org/sections/health-shots/2020/07/01/886299190/it-does-not-have-to-be-100-000-cases-a-day-fauci-urges-u-s-to-follow-guidelines>
- Brickman P., D. Coates and R. Janoff-Bulman (1978), 'Lottery winners and accident victims: is happiness relative?', *Journal of Personality and Social Psychology*, **36**(8): 917.
- Brown T. B., B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell et al. (2020), 'Language models are few-shot learners', arXiv Preprint arXiv:2005.14165.
- Bubeck S., V. Chandrasekaran, R. Eldan, J. Gehrke, E. Horvitz, E. Kamar, P. Lee, Y. T. Lee, Y. Li, S. Lundberg et al. (2023), 'Sparks of artificial general intelligence: early experiments with gpt-4', arXiv Preprint arXiv:2303.12712.
- Callaway E. (2020), "'It will change everything": DeepMind's AI makes gigantic leap in solving protein structures', *Nature*, **588**(7837): 203–205.
- Camerer C., S. Issacharoff, G. Loewenstein, T. O'Donoghue and M. Rabin (2003), 'Regulation for conservatives: behavioral economics and the case for "asymmetric paternalism"', *University of Pennsylvania Law Review*, **151**(3): 1211–1254.
- Chater N. and G. Loewenstein (2023), 'The i-frame and the s-frame: how focusing on individual-level solutions has led behavioral public policy astray', *Behavioral and Brain Sciences*, **46**: e147.
- Chater N. and M. Oaksford (2013), 'Programs as causal models: speculations on mental programs and mental representation', *Cognitive Science*, **37**: 1171–1191.
- Cherry T. L., S. Kallbekken and S. Kroll (2012), 'The acceptability of efficiency-enhancing environmental taxes, subsidies and regulation: an experimental investigation', *Environmental Science & Policy*, **16**: 90–96.

- Choi J. J., D. Laibson, J. Cammarota, R. Lombardo and J. Beshears (2023). 'Do automatic savings policies actually increase savings?' Working Paper.
- Cialdini R. B. and N. J. Goldstein (2004), 'Social influence: compliance and conformity', *Annual Review of Psychology*, **55**: 591–621.
- Coombe S. (2023). 'How does the UK get to a fully functional EPR?' <https://resource.co/article/how-does-uk-get-fully-functional-epr> (20 October 2023).
- Currie J., L. Davis, M. Greenstone and R. Walker (2015), 'Environmental health risks and housing values: evidence from 1,600 toxic plant openings and closings', *American Economic Review*, **105**(2): 678–709.
- Daron Acemoglu, Simon Johnson, and James A. Robinson. 2001. "The Colonial Origins of Comparative Development: An Empirical Investigation." *American Economic Review*, 91 (5): 1369–1401.
- Deaton A. (2020), 'Randomization in the tropics revisited: a theme and eleven variations', in F. Bédécarrats, I. Guérin and F. Roubaud (eds), *Randomized Controlled Trials in the Field of Development: A Critical Perspective*, Oxford University Press, Revised, 29–46.
- Deaton A. and N. Cartwright (2018), 'Understanding and misunderstanding randomized controlled trials', *Social Science & Medicine*, **210**: 2–21.
- DellaVigna S., W. Kim and E. Linos (2022), *Bottlenecks for Evidence Adoption* (No. w30144), National Bureau of Economic Research.
- DellaVigna S. and E. Linos (2022), 'RCTs to scale: comprehensive evidence from two nudge units', *Econometrica*, **90**(1): 81–116.
- Deschênes O., M. Greenstone and J. S. Shapiro (2017), 'Defensive investments and the demand for air quality: evidence from the NOx budget program', *American Economic Review*, **107**(10): 2958–2989.
- Dessart L. and V. Pitardi (2019), 'How stories generate consumer engagement: an exploratory study', *Journal of Business Research*, **104**: 183–195.
- Diepeveen S., T. Ling, M. Suhrcke, M. Roland and T. M. Marteau (2013), 'Public acceptability of government intervention to change health-related behaviours: a systematic review and narrative synthesis', *BMC Public Health*, **13**(1): 1–11.
- Dobbie W., J. Goldin and C. S. Yang (2018), 'The effects of pre-trial detention on conviction, future crime, and employment: evidence from randomly assigned judges', *American Economic Review*, **108**(2): 201–240.
- Duflo E., Kremer M. (2005). Use of randomization in the evaluation of development effectiveness. In Feinstein O., Ingram G. K., Pitman G. K. (Eds.), *Evaluating development effectiveness* (pp. 205–232). London, England: Transaction.
- Fagan B. (2019), *The Little Ice Age: How Climate Made History 1300-1850*, London: Hachette.
- Fhanér G. and M. Hane (1979), 'Seat belts: opinion effects of law-induced use', *Journal of Applied Psychology*, **64**(2): 205–212.
- Fishbane A., A. Ouss and A. K. Shah (2020), 'Behavioral nudges reduce failure to appear for court', *Science*, **370**(6517): eabb6591.
- Frederick S., and G. Loewenstein (1999), '16 Hedonic adaptation', in D. Kahneman ED, and N. Schwarz (eds New York: Russell Sage Foundation), *Well-Being the Foundations of Hedonic Psychology*, 302–329.
- Frieden T. R. (2017), 'Evidence for health decision making — beyond randomized, controlled trials', *New England Journal of Medicine*, **377**: 465–475.
- Golman R., G. Loewenstein, K. O. Moene and L. Zarri (2016), 'The preference for belief consonance', *Journal of Economic Perspectives*, **30**(3): 165–188.
- Gopnik A., C. Glymour, D. M. Sobel, L. E. Schulz, T. Kushnir and D. Danks (2004), 'A theory of causal learning in children: causal maps and bayes nets', *Psychological Review*, **111**(1): 3.
- Goss S. (2023, July 12), 'Letter to representative John Larson'. https://www.ssa.gov/OACT/solvency/JLarson_20230712.pdf
- Green D. P. and D. Winik (2010), 'Using random judge assignments to estimate the effects of incarceration and probation on recidivism among drug offenders', *Criminology*, **48**(2): 357–387.
- Griffiths T. L., N. Chater, C. Kemp, A. Perfors and J. B. Tenenbaum (2010), 'Probabilistic models of cognition: exploring representations and inductive biases', *Trends in Cognitive Sciences*, **14**(8): 357–364.
- Hagmann D., E. H. Ho and G. Loewenstein (2019), 'Nudging out support for a carbon tax', *Nature Climate Change*, **9**(6): 484–489.
- Hagmann D., Y. T. Liao, N. Chater and G. Loewenstein (2023), 'Costly distractions: focusing on individual behavior undermines support for systemic reforms', *Available at SSRN*.

- Haidt J. (2008), 'Morality', *Perspectives on Psychological Science*, 3(1): 65–72.
- Hargreaves Heap S. P. (2024), 'What to do about preference change? Lessons from "philosophy, politics and economics"', *Review of Behavioral Economics*, 11(2): 275–294.
- Hausmann R. and D. Rodrik (2003), 'Economic development as self-discovery', *Journal of Development Economics*, 72(2): 603–633.
- Haynes L., O. Service, B. Goldacre and D. Torgerson (2012), 'Test, Learn, Adapt: Developing Public Policy with Randomized Controlled Trials', UK Behavioral Insights Team.
- Hinton G. E. and R. R. Salakhutdinov (2006), 'Reducing the dimensionality of data with neural networks', *Science*, 313(5786): 504–507.
- Infante G., G. Lecouteux and R. Sugden (2016), 'Preference purification and the inner rational agent: a critique of the conventional wisdom of behavioural welfare economics', *Journal of Economic Methodology*, 23(1): 1–25.
- Irzik G. (1985), 'Popper's piecemeal engineering: what is good for science is not always good for society', *The British Journal for the Philosophy of Science*, 36(1): 1–10.
- Iwry J. M. and D. C. John (2021, February 12), 'The Automatic IRA at 15: helping Americans build retirement security'. Brookings. <https://www.brookings.edu/articles/the-automatic-ira-at-15-helping-americans-build-retirement-security/>
- Jackson M. (2023), 'The social sciences are increasingly ill-equipped to design system-level reforms', *Behavioral and Brain Sciences*, 46: E162.
- Janusch N., S. Kroll, C. Goemans, T. L. Cherry and S. Kallbekken (2021), 'Learning to accept welfare-enhancing policies: an experimental investigation of congestion pricing', *Experimental Economics*, 24: 59–86.
- Johnson S. G., A. Bilovich and D. Tuckett (2023), 'Conviction narrative theory: a theory of choice under radical uncertainty', *Behavioral and Brain Sciences*, 46: e82.
- Kahneman D. and S. Frederick (2002), 'Representativeness revisited: attribute substitution in intuitive judgment: the psychology of intuitive judgment', in T. Gilovich, D. Griffin and D. Kahneman (eds), *Heuristics and Biases*, New York, NY: Cambridge University Press, 49–81.
- Klandermans P. G. (2014), 'Identity politics and politicized identities: identity processes and the dynamics of protest', *Political Psychology*, 35(1): 1–22.
- Koriat A., L. Sheffer and H. Ma'ayan (2002), 'Comparing objective and subjective learning curves: judgments of learning exhibit increased underconfidence with practice', *Journal of Experimental Psychology: General*, 131(2): 147.
- Laibson D. (2020, January 3), 'Nudges Are Not Enough'. American Economic Association. <https://www.aeaweb.org/conference/webcasts/2020>
- Lakoff G. (2010), 'Why it matters how we frame the environment', *Environmental Communication*, 4(1): 70–81.
- Loewenstein G. and P. A. Ubel (2008), 'Hedonic adaptation and the role of decision and experience utility in public policy', *Journal of Public Economics*, 92(8-9): 1795–1810.
- Luca M. and M. H. Bazerman (2021), *The Power of Experiments: Decision Making in a Data-driven World*, Cambridge, MA: MIT Press.
- Mabsout R. (2022), 'John Stuart Mill, soft paternalist', *Social Choice and Welfare*, 58(1): 161–186.
- March J. G., and J. P. Olsen (2008), 'The logic of appropriateness', in R. E. Goodin, M. Moran and M. Rein (eds), *The Oxford Handbook of Public Policy*, Oxford: Oxford University Press 689–708.
- Marron D. B. and A. C. Morris (2016), 'How to use carbon tax revenues', *SSRN Electronic Journal*.
- Mulgan G. (2022), *Another World Is Possible: How to Reignite Social and Political Imagination*, London, UK: Hurst Publishers.
- Murthukrishna M., J. Henrich and E. Slingerland (2021), 'Psychology as a historical science', *Annual Review of Psychology*, 72: 717–749.
- Newell A. (1973), 'You can't play twenty questions with nature and win', in W. G. Chase (ed), *Visual Information Processing*, New York, NY: Academic Press.
- North D. C. (1990), *Institutions, Institutional Change and Economic Performance*, Cambridge: Cambridge University Press.
- North D. C. and R. P. Thomas (1973), *The Rise of the Western World: A New Economic History*, Cambridge: Cambridge University Press.

- Oreskes N. and E. M. Conway (2023), *The Big Myth: How American Business Taught Us to Loathe Government and Love the Free Market*, New York, NY: Bloomsbury Publishing USA.
- Parker G. (2013), *Global Crisis: War, Climate Change, and Catastrophe in the Seventeenth Century*, New Haven, CT: Yale University Press.
- Pearl J. (2009), *Causality*, Cambridge: Cambridge University Press.
- Peri G. and V. Yassenov (2019), 'The labor market effects of a refugee wave: synthetic control method meets the Mariel boatlift', *Journal of Human Resources*, **54**(2): 267–309.
- Pezzullo P. and R. Cox (2022), *Environmental Communication and the Public Sphere*, 6th edn, Los Angeles, CA: Sage.
- Popper K. R. (1957), *The Poverty of Historicism*, Boston: Beacon Press.
- Reed M., P. O'Reilly and J. Hall (2019), 'The economics and politics of carbon taxes and regulations: evidence from voting on Washington State's Initiative 732', *Sustainability*, **11**(13): 3667.
- Rescorla R. A. and A. R. Wagner (1972), 'A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement', in A. H. Black and W. F. Prokasy (eds), *Classical Conditioning II: Current Research and Theory*, New York: Appleton-Century-Crofts, 64–99.
- Rizzo M. J. (1980), 'Law amid flux: the economics of negligence and strict liability in tort', *The Journal of Legal Studies*, **9**(2): 291–318.
- Rizzo M. J. (2021), 'Are general inflexible rules stable? An examination of David Hume's jurisprudence', *Journal of Economic Behavior & Organization*, **182**(C): 29–38.
- Rizzo M. J. and D. G. Whitman (2009), 'The knowledge problem of new paternalism', *Brigham Young University Law Review*, **2009**(4): 905–968.
- Rizzo M. J. and G. Whitman (2018), 'Rationality as a process', *Review of Behavioral Economics*, **5**(3-4): 201–219.
- Rizzo M. J., and G. Whitman (2019), *Escaping Paternalism: Rationality, Behavioral Economics, and Public Policy*, New York, NY: Cambridge University Press.
- Rodrik D. (2009), 'The new development economics: We shall experiment, but how shall we learn?', in J. Cohen and W. Easterly (ed.), *What Works in Development? Thinking Big and Thinking Small*, Washington, DC: Brookings Institution Press. Chap. 2.
- Roser M., E. Ortiz-Ospina and H. Ritchie (2013), 'Life Expectancy', *Our World in Data*. <https://ourworldindata.org/life-expectancy>
- Saccardo S., H. Dai, M. Han, N. Raja, S. Vangala and D. Croymans (2022), 'Assessing nudge scalability: two lessons from large-scale RCTs', (March 5, 2022) Available at SSRN: <https://ssrn.com/abstract=3971192>
- Scheffer M., I. Van De Leemput, E. Weinans and J. Bollen (2021), 'The rise and fall of rationality in language', *Proceedings of the National Academy of Sciences*, **118**(51): e2107848118.
- Schrittwieser J., I. Antonoglou, T. Hubert, K. Simonyan, L. Sifre, S. Schmitt and D. Silver (2020), 'Mastering Atari, go, chess and shogi by planning with a learned model', *Nature*, **588**(7839): 604–609.
- Seligman M. (2023), 'Psychological history and predicting the future', *Possibility Studies & Society*, **1**(1–2): 206–210.
- Singal J. (2013, July 11), 'Daniel Kahneman's gripe with Behavioral Economics', *The Daily Beast*. <https://www.thedailybeast.com/daniel-kahnemans-gripe-with-behavioral-economics>
- Solnit R. (2021, August 23), 'Big oil coined "carbon footprints" to blame us for their greed. Keep them on the hook | rebecca Solnit', *The Guardian*. <https://www.theguardian.com/commentisfree/2021/aug/23/big-oil-coined-carbon-footprints-to-blame-us-for-their-greed-keep-them-on-the-hook>
- Soni S., L. F. Klein and J. Eisenstein (2021), 'Abolitionist networks: modeling language change in Nineteenth-Century Activist Newspapers', *Journal of Cultural Analytics*, **6**(1).
- Sugden R. (2018), *The Community of Advantage: A Behavioral Economist's Defence of the Market*, Oxford: Oxford University Press.
- Sunstein C. R. and R. H. Thaler (2003), 'Libertarian paternalism is not an oxymoron', *The University of Chicago Law Review*, **70**(4): 1159–1202.
- Sutton R. S., and A. G. Barto (2018), *Reinforcement Learning: An Introduction (2nd Edition)*, Cambridge, MA: MIT Press.
- Thaler R. (2023), 'Nudging is being framed', *Behavioral and Brain Sciences*, **46**: E177.
- Thaler R. H. and C. R. Sunstein (2008), *Nudge: Improving Decisions about Health, Wealth, and Happiness*, New Haven, CT: Yale University Press.

- Thaler R. H., C. R. Sunstein and J. P. Balz (2014), 'Choice architecture', *The Behavioral Foundations of Public Policy*.
- Thunström L. (2019), 'Welfare effects of nudges: the emotional tax of calorie menu labeling', *Judgment and Decision Making*, **14**(1): 11–25.
- Thunström L., B. Gilbert and C. J. Ritten (2018), 'Nudges that hurt those already hurting – distributional and unintended effects of salience nudges', *Journal of Economic Behavior & Organization*, **153**: 267–282.
- Tilly C. (2004), *Social Movements, 1768–2012*, Boulder, CO: Paradigm Publishers.
- Tor A. and J. Klick (2022), 'When should governments invest more in nudging? Revisiting Benartzi et al. (2017)', *Review of Law & Economics*, **18**(3): 347–376.
- Ubel P. A., G. Loewenstein and C. Jepson (2005), 'Disability and sunshine: can hedonic predictions be improved by drawing attention to focusing illusions or emotional adaptation?', *Journal of Experimental Psychology: Applied*, **11**(2): 111.
- VanEpps E. M., A. Molnar, J. S. Downs and G. Loewenstein (2021), 'Choosing the light meal: real-time aggregation of calorie information reduces meal calories', *Journal of Marketing Research*, **58**(5): 948–967.
- Vong W. K., W. Wang, A. E. Orhan and B. Lake (2024) 'Grounded language acquisition through the eyes and ears of a single child', *Science*, **383**: 504–511.
- Walker E. T. (2014), *Grassroots for Hire: Public Affairs Consultants in American Democracy*, Cambridge, UK: Cambridge University Press.
- Werfel S. H. (2017), 'Household behaviour crowds out support for climate change policy when sufficient progress is perceived', *Nature Climate Change*, **7**(7): 512–515.
- Willis L. E. (2008), 'Against financial-literacy education', *Iowa Law Review*, **94**: 197.
- Wills K. (2023), 'The anti-Ulez protests: a lightning rod for rage over London's ultra-low emission zone expansion'. *Evening Standard* (13 June 2023). <https://www.standard.co.uk/insider/anti-ulez-protests-ultra-low-emissions-zone-drivers-rage-sadiq-khan-policy-b1087402.html>
- Wing C., K. Simon and R. A. Bello-Gomez (2024), 'Designing difference in difference studies: best practices for public health policy research', *Annual Review of Public Health.*, **39**: 453–469.
- Woodside A. G., S. Sood and K. E. Miller (2008), 'When consumers and brands talk: storytelling theory and research in psychology and marketing', *Psychology & Marketing*, **25**(2): 97–145.
- Yang Y. and S. T. Piantadosi (2022), 'One model for the learning of language', *Proceedings of the National Academy of Sciences*, **119**(5): e2021865119.