Ana Tanasoca* and Andrew Leigh

The Democratic Virtues of Randomized Trials

https://doi.org/10.1515/mopp-2022-0039
Published online February 17, 2023

Abstract: Democratic alternation in power involves uncontrolled policy experiments. One party is elected on one policy platform that it then implements. Things may go well or badly. When another party is elected in its place, it implements a different policy. In imposing policies on the whole community, parties in effect conduct non-randomized trials without control groups. In this paper, we endorse the general idea of policy experimentation but we also argue that it can be done better by deploying in policymaking randomized controlled trials. We focus primarily on the democratic benefits of using randomized trials in policymaking and on how they can enhance the democratic legitimacy of policy. We argue that randomized trials resonate well with three key democratic principles: non-arbitrariness, revisability and public justification. Randomized trials' contribution to non-arbitrariness and revisability is not unique; other types of evidence can advance these democratic principles as well. But through their peculiar democratic scrutability, randomized trials are well-equipped to contribute to the public justifiability of policy.

Keywords: randomized trials, public policy, democratic legitimacy, public justification

1 Introduction: Policy as Experiment

Democratic alternation in power involves, by its very nature, uncontrolled policy experiments (Campbell 1969; Rivlin 1971, ch. 5).[1] One party is elected on one policy platform that it then implements. Things may go well or badly. When another party is

1 Dewey (2012) similarly conceives democracy as entailing continuous experimentation towards problem-solving and learning from past experience.

Ana Tanasoca and Andrew Leigh contributed equally to this work and author ordering was randomised using the American Economic Association's Author Randomisation Tool.

***Corresponding author: Ana Tanasoca**, Department of Philosophy, Macquarie University, North Ryde NSW 2109, Sydney, Australia, E-mail: ana.tanasoca@mq.edu.au
Andrew Leigh, Parliament House, Canberra ACT 2600, Australia, E-mail: andrew.leigh.mp@aph.gov.au. https://orcid.org/0000-0002-5639-0509

elected in its place, it implements its own different policy. Policy experimentation is thus inevitable: in imposing policies on the entire community, parties in effect conduct non-randomized trials without control groups.[2] That is questionable on epistemic grounds: we cannot know what would have happened otherwise. In this paper, we endorse the general idea of policy experimentation and we show how it is closely tied to democratic governance.[3] But we also argue, more specifically, that policy experimentation can be done better by deploying in policymaking the same kind of randomized controlled trials that have become ubiquitous in medical research and have recently helped develop vaccines to fight the Covid-19 pandemic.[4]

The experimental approach relies on randomized trials to test the impact of policy. Non-experimental (or econometric) policy evaluation relies instead on statistical analysis of behavioral data. Here we put forward a case for using well-conducted[5] randomized trials to test and choose policy, in addition to other types of evidence and knowledge that may be relevant (to justify conducting the trial, figuring out what treatment we should test, and how we should set up the test). In contrast to existing debates in the philosophy of science and economics that have focused largely on debating the scientific[6] and ethical[7]

2 That is more justifiable in some cases than in others. For example, in the case of emergencies created by 'black swan' events (Taleb 2007), governments will have to act very quickly and on the basis of little evidence. They might have to adopt some ad-hoc measures on the basis of theory and extrapolation from existing evidence.

3 Such policy experiments are a feature of the piecemeal social engineering defended by Popper (1980, 162) as an alternative to utopian engineering.

4 Political systems have occasionally used randomness to select rulers, such as sortition in ancient Athens (Ober 2017). Some advocate random selection of legislators as an efficient and effective form of governance today (see e.g. Landemore 2013).

5 'Well-conducted' because the reliability of any type of evidence (whether experimental or non-experimental) ultimately depends on the power of the methodology, the quality of the research design and the competence of the researcher. As we clarify in Section 3, 'well-conducted' randomized trials are trials that use sufficiently large and representative samples, with blinding where practicable. Furthermore, in such trials researchers pay attention to and try to limit the influence of other confounding factors that may affect the results of the trial (see our discussion in Section 3). Last, but not least, randomized trials are well-conducted insofar as they abide by robust ethical protocols.

6 The main question at the heart of many of those debates is whether or not randomized trials yield causal inferences and estimates of average treatment effects that are more reliable or credible than those of other empirical methods (e.g. Deaton and Cartwright 2018; for a reply see Backmann 2017). For the purposes of our argument, it is not necessary to take a position on where randomized trials fit in the epistemic pecking order of evidence. Where other high-quality evidence exists, it should be considered in conjunction with evidence from randomized trials. Despite critics claiming the contrary, this is the official stance of the Evidence Based Medicine Working Group (1992, 2421).

7 Some have argued that randomized trials are morally problematic because in offering estimates of average effects they encourage utilitarian thinking. Utilitarianism in public policy is certainly not without defenders (see e.g. Goodin 1995), although we do not here advocate any particular moral theory. We note that randomized trials yield information about the distribution of effects as well as

strengths and weaknesses of this method and its design, here we focus primarily on the democratic benefits of using randomized trials in policymaking—that is, on how they can enhance the democratic legitimacy of policy. We argue, specifically, that randomized trials resonate well with three key democratic principles: non-arbitrariness, revisability, and public justification. Randomized trials' contribution to non-arbitrariness and revisability is not unique; other types of evidence can advance these democratic principles as well. Randomized trials are, however, uniquely well-equipped to contribute to the public justifiability of policy through their peculiar democratic scrutability.

The relationships between democracy and randomized trials are thus threefold.

- First, both democracy and randomized trials contribute, in different ways, to the non-arbitrariness of political rule. Democratic accountability makes government policy systematically responsive to the wishes of the people, rather than some arbitrary fiat imposed from above. Well-conducted randomized trials, coupled with other good-quality non-experimental data, help make policy systematically responsive to the best evidence about the effectiveness of government programs, rather than to the whims or prejudices of policymakers. This is the first sense in which democracy and randomized trials work in unison.
- Second, randomized trials take advantage of democracy's commitment to revisability. The inherent revisability of democratic decisions makes democracy the sort of regime that can rely on randomized trials in policymaking.
- Third, in contrast to other types of evidence, evidence from randomized trials has greater democratic scrutability. Randomized trials are thus better able to provide public justification for policy to the wider public, which should make democracies want to rely more on randomized trials to increase the legitimacy of their decisions. While econometric analyses also serve the goals of non-arbitrariness and revisability, they are more epistemically opaque to the general public and thus less able to provide public justification.

The fact that well-conducted randomized trials can readily satisfy all three of those democratic principles gives us good democratic reasons to seek out such evidence for policymaking. Some of those same (e.g., epistemic) considerations may also recommend randomized trials to non-democrats of certain sorts (e.g., epistocrats, enlightened despots). The claim of this paper is not that the arguments for randomized trials imply democracy or that randomized trials are 'inherently' democratic, but rather (roughly speaking) that arguments for democracy recommend randomized trials because of the ways in which such trials can enhance the legitimacy of democratic outputs produced by the democratic

---

the average; and effects on subgroups of the population of special concern can be separated out. So there is no reason to think that randomized trials should appeal only to utilitarians.

electoral system. Indeed, as shown below, empirical evidence lends support to this argument: the more democratic a regime is, the more likely it is to conduct randomized trials.

At least from a democratic legitimacy perspective, there may be benefits that randomized trials more easily secure than other types of evidence. This however does not necessarily make randomized trials, all things considered, superior to other types of evidence. Rather, the arguments here should be taken merely as *pro tanto* reasons for using randomized trials in policymaking alongside other types of evidence and knowledge. The aim of this paper is to provide a partial defense, from a politico-democratic perspective, of using randomized trials more widely in policymaking. To make our case about the democratic virtues of using such evidence, we start from the modest premise that, when well-conducted and coupled with existing knowledge, randomized trials do have independent epistemic value and in virtue of that they can help us choose better policy.[8] Randomized trials cannot answer all policy questions and, even where they are applicable, they should not necessarily have the last word. In the next section we explain the epistemic merits and limits of such evidence. With those caveats in mind, however, we do believe that randomized trials are particularly well-suited at simultaneously serving the three democratic ideals discussed here: non-arbitrariness, revisability and, most particularly, public justification.

We start by discussing the place of randomized trials in different political regimes (Section 2). We follow by explaining the methodology of randomized trials, and its epistemic merits and limitations (Section 3). We then discuss the democratic appeal of randomized trials and how they can help promote non-arbitrariness, revisability and public justification (Section 4). We next propose a new way in which deliberative participatory processes might be conducted as a means of determining which policies should be tested through large, randomized trials (Section 5). The final section concludes.

2 Randomized Trials in Alternative Political Regimes

The central argument of this article is that some of the virtues of randomized trials are particularly (perhaps uniquely) suited to electoral democracy. Randomized trials can enhance the non-arbitrariness of political rule—a value that is at stake in a

8 Indeed, even randomized trials' most fierce critics acknowledge as much (see Deaton and Cartwright 2018, 3).

democracy, but not in a dictatorship. Randomized trials also dovetail with the revisability that is naturally embedded into electoral systems, whose raison d'être is to enable the change of both political personnel and policy. Using randomized trials to inform policymaking is just a natural extension of the exploratory and experimental character of liberal democracies that is already enshrined in their reliance on periodic elections as a mechanism for filling office.

As we have said, a less democratic regime might in principle benefit from using randomized trials in policymaking, insofar as that would enhance the actual performance of the regime and provide better public justification for its policies. However, as we will later see, autocracies and countries where elections are less competitive tend to conduct fewer randomized trials. This may be because undemocratic regimes rely on strategies such as bargaining with elites and the use of force to maintain power, and are therefore less likely to need rigorous scientific evidence of the positive impact of their policies. There remains the question of how the increased use of randomized trials to select policy would impact democracy more generally, and electoral campaigns in particular. If a government's policy choices will be at least partly determined by the results of randomized trials, that will surely restrict what campaign promises can be made and the extent to which voters can select among parties on the basis of their promises. In the extreme case, voters might find themselves choosing not between parties that promise to implement different policies, but between parties that promise to trial different policies.

These are reasonable concerns, but they are not fatal to our proposal. First, parties could still promise in advance to pursue specific policies whose effectiveness has already been confirmed by existing randomized controlled trials. Second, even if new randomized controlled trials would need to be organized, that would not prevent campaign promises. Parties could still make promises about the goals they would achieve when in government, leaving the means of achieving those goals to be determined by subsequent randomized trials (e.g. Campbell 1998).

On reflection, this may not be as different from the current democratic system as it appears on first blush. When it comes to pharmaceutical treatments for cancer, for example, democratic leaders are more likely to promise support for clinical trials of a range of treatments than they are to back a particular experimental treatment. More generally, no party in a democratic system provides an exhaustive itemization of their policies. Democratic accountability is essentially about providing an acceptable retrospective accounting, come the next election, for what the government has done in its last term in office (Fiorina 1981; Key 1966; Schumpeter 1950). And that accounting should turn principally on accomplishments, not intentions. Consider a party that won government on the basis of a very specific set of policy proposals that led to economic disaster. They might try standing for re-election saying, 'At least we did precisely what we promised we would do!' But they are not likely to win many votes on that basis, nor

should they. Democratic accountability judges governments on promised outcomes (implicitly in this case, economic prosperity) rather than promised instruments (explicitly in this case, pursuing one economic policy rather than another).

3 Epistemic Benefits of Randomized Trials

In 1953, the University of Pittsburgh's Jonas Salk was confident that he had developed an effective vaccine for polio. The previous year, over 50,000 people in the United States had contracted polio, and more than 3000 had died from the disease (Schmeck 1995). Yet other experts were skeptical. Salk's critics included Nobel laureate John Enders, and virologist Albert Sabin, who would go on to develop his own vaccine five years later (Meldrum 1998).

To determine the efficacy of the Salk vaccine, the United States embarked on one of the largest experiments in history. Over 400,000 US children were randomly assigned to be given either the Salk vaccine or an injection of saline solution. On April 12, 1955, the trial was declared to be a success. Mass vaccination of all US school children began, and the polio epidemic came to an end. In 2020, there were fewer than 200 cases of wild polio worldwide, and many hopes that it might ultimately be eradicated, as smallpox was in 1980.

The polio randomized trials were part of a shift in medical research from non-experimental to experimental approaches. While randomized trials remain the main method for testing and trialing new drugs and treatments, they are also widely used in other disciplines such as development economics.[9] They have also become increasingly prevalent as a tool for evaluating policies in education, crime prevention, social policy, employment, philanthropy and many more policy fields. Several governments have created behavioral economics units to conduct randomized trials of new and existing programs.[10]

All good-quality evidence and knowledge has its place in medicine, economics, and—of course—public policymaking (Global Commission on Evidence to Address

9 One study found that from 2003 to 2012 there was a tenfold increase in the annual number of randomized trials in development economics, a tenfold increase in education, and a fivefold increase in the field of social work (White 2019). Another study found that the volume of randomized trials published in well-regarded medical journals was 10 times higher in the 2010s than the 1990s (Vinkers et al. 2021).

10 Many recent randomized policy trials evaluate behavioral economics interventions (dubbed 'nudging' by Thaler and Sunstein 2009). Such interventions include cell phone text message reminders, messages that emphasize how most other people behave, incentives that are framed so as to trigger loss aversion, and interventions that exploit present bias by asking people to make a virtuous commitment that binds their future self.

Societal Challenges 2022). So what distinguishes evidence from randomized trials from other types of evidence and what are its epistemic strength and limits?

Randomized trials are based on the statistical fact that if participants are assigned to treatment and control groups based on chance then, with a sufficiently large sample size, the two groups will enter the experiment statistically identical to one another. The two groups will, on average, share the same observable characteristics (e.g., age, height, income) and the same unobservable characteristics (e.g., grit, friendliness, genetic predisposition to disease). With a sufficiently large sample, observed differences will be due only to the treatment. By randomly allocating subjects to treatment and control groups, researchers avoid the possibility of selection bias: that those who sign up for a new treatment may be different (or may be on a different trajectory) than those who do not sign up.

Methodologically, researchers conducting a randomized trial begin with a preferred theory, but must be impartial in their evaluation of the theory. The increasing practice of researchers lodging pre-analysis plans setting out how they plan to conduct the experiment and analyze the results makes it more difficult for researchers to cherry-pick results (by, for example, changing their outcome of interest, or focusing on sub-samples). As in medicine, it is becoming increasingly difficult for social scientists to publish the results of randomized experiments in leading journals if they failed to lodge a pre-analysis plan (AEA 2021; ICMJE 2021). With smaller samples, there is a slight efficiency gain from stratified randomization, in which the population is first divided into subgroups with similar observable characteristics, and these subgroups are then randomized into treatment and control groups. With either simple randomization or stratified randomization, it is common practice for researchers to report the observable characteristics at baseline for the treatment and control groups to reassure readers that the randomization has been carefully conducted and that any observed differences after the intervention can be attributed to the treatment.

The signal advantage of randomization lies in determining the counterfactual —what would have occurred in the absence of the intervention. By definition, the counterfactual represents a state of the world that ordinarily would not be observed. Yet it is crucial to determining the impact of a policy. Most sick people who go to a doctor would have eventually recovered anyway. To attribute the full extent of the recovery to the doctor would therefore be to give the doctor too much credit. Similarly, a policy evaluation that merely follows the same people over time risks overestimating the impact of the intervention. For a community that is hit by a natural disaster, the correct counterfactual is not that they would have stayed devastated forever. For children in a classroom, the right counterfactual is not that they would have stayed at a static level of achievement. For jobseekers, the proper counterfactual is not that they would have stayed unemployed for the remainder of their life.

To address this problem, many quasi-experimental econometric techniques have been developed. Regression analysis attempts to hold constant the observable characteristics of treatment and control populations. Matching estimators are similar, but allow for non-linearities. Differences-in-differences follows the trajectory of non-randomly treated and untreated groups of people. Regression discontinuity compares policy impacts at a clearly defined break-point such as a test score cut-off. Instrumental variables use variation from an external source (such as unexpected weather events) which affects the outcome variable only by affecting the independent variable of interest.

Such quasi-experimental approaches represent important methodological advances over simple before–after studies. While quasi-experimental findings should be taken into account by policymakers, the fact remains that they can still produce biased results. One problem is that of selection bias. If a job training program disproportionately attracts highly motivated participants, then a comparison between those who choose to sign up and those who do not choose to sign up may overstate the true effect of the intervention.[11] A large literature has compared the results of experimental (randomized) and non-experimental evaluations, frequently finding that quasi-experimental evaluation approaches, including differences-in-differences and regression discontinuity, do not always match the benchmark results from well-conducted randomized trials. For example, a randomized evaluation of the impact of migration found that most non-experimental methods overstated the gains from migration, some by as much as 80 percent (McKenzie, Stillman, and Gibson 2010). A study of the impact of job training found that non-experimental methods produced estimates around twice the true effect—as estimated from a randomized evaluation (Sauermann and Stenberg 2020). A study of agricultural technology adoption in a developing country context found that an instrumental variable approach would have overestimated the actual impact, while other non-experimental approaches would have underestimated the true effect of the program (Omotilewa and Ricker-Gilbert

11 Naturally, these are not the only factors that might lead the results of an evaluation to differ from the true impact of an intervention implemented at scale. Other problems can occur if the evaluation is based on a group of people who are unrepresentative of the broader population, implemented by people who are not representative of those who would be running the program at scale, or ended before the effects of the program are fully realized. The evaluation will be biased if attrition is correlated with whether someone is in the treatment or control group. Moreover, small-scale programs may not capture general equilibrium effects, such as the possibility that a job training program helps those in the program, but only at the expense of those outside the program (and therefore has no overall impact on the unemployment rate). See, for example, Moffitt (1992), Garfinkel et al. (1992) and Burtless (1995) for a discussion of these issues. We do not address them in detail here because they largely apply to both randomized evaluations and non-randomized evaluations alike.

2019). An experiment on the impact of hybrid university classes found that non-experimental techniques would have badly misestimated the impact, due to selection bias (Joyce et al. 2014).

Other comparisons of randomized and quasi-experimental methods have found that researcher discretion plays a larger role in non-randomized studies. Natural experiment techniques allow more scope for researcher discretion in choosing the appropriate comparison groups, as well as in deciding upon the precise methodology for making the comparison. This can have a significant impact on the results (Bloom et al. 2002; Gleason, Resch, and Berk 2012; St. Clair, Cook, and Hallberg 2014). These are not arguments for abandoning non-experimental or quasi-experimental approaches. They are merely arguments for policymakers to supplement such evidence with evidence from randomized trials. We stress supplementing because policymakers must be aware of both strengths and limitations of each sort of evidence. They should be careful how they use the results of randomized trials just as they should be with all other sorts of evidence.

As with any form of evaluation, the results of randomized evaluations will be context-sensitive, shaped by culture and institutions, as well as the specific way in which the program is designed. This is as true of a simple before–after evaluation as it is of the most rigorously conducted randomized trial. Even the best evaluation measures the program as delivered, which may be different from the way it would be implemented at scale. To the extent that interventions are affected by factors such as the ethnicity of the participants, the social setting of the intervention or the historical period in which they are delivered, evaluations may not extrapolate perfectly into different contexts. Such issues also affect systematic reviews, which must take care when aggregating results from multiple evaluations.[12]

The best-conducted randomized trials use large, representative samples. Even so, because participation in many randomized trials is voluntary, the results may not perfectly generalize to the entire population. The trial results will, nonetheless,

12 Economists Paul Glewwe and Karthik Muralidharan (2016) illustrate the point by reference to four studies that found students in developing countries who were randomly assigned to receive textbooks did no better on standardized tests than students without textbooks. A naïve response to reading these studies might have been to simply conclude 'textbooks do not matter'. But as Glewwe and Muralidharan point out, it is important to unpack the findings. In the Sierra Leone evaluation, schools put the textbooks in storage rather than delivering them to the classrooms. In a study conducted in India, free textbooks led parents to reduce the amount that they spent on their children's education. An assessment of textbook distribution in Tanzania found that teachers did not have an incentive to incorporate textbooks into their teaching. And a study in Kenya found that textbooks did help the top students, but not the remainder, who were unable to read. Understanding the context of the studies helps researchers and policymakers form a fuller picture of the intervention.

provide an unbiased estimate of the impact of the treatment on those who do choose to participate. Moreover, many policies will themselves require citizens' voluntary participation. At least in those cases, self-selection would not necessarily provide a misleading picture of the policy impact—quite the contrary, it would ensure that the policy is tested on the same population that would be affected by that policy. Yet, as with any type of evidence, careful judgment should be exercised to determine how much weight, depending on the context and the question we are trying to answer, we should place on the results of any given randomized trial. The best randomized trials will be those with fairly large and representative samples that have (where possible) been properly blinded[13]—and our argument here applies most strongly to those types of trials.

Randomized trials can be valuable as a way of 'proving concept' and in situations where there is little existing evidence as to the efficacy of any given treatment or policy (Deaton and Cartwright 2018, 5, 13). Perhaps the most appealing advantage of randomized trials is that they provide unbiased estimates. As we will see later on in Section 4, randomized trials' unbiasedness is valuable not just as an epistemic good; it also makes such evidence particularly valuable from a democratic perspective. Combining evidence from randomized trials (which are relatively rare) with evidence from other forms of evaluation (which are more commonly used in policymaking) is likely to produce results that are more precise and less biased than if policymakers eschew randomized trials entirely.

Needless to say, as with any type of data, the results of any given randomized trial should also not be blindly trusted or acted upon. It is equally important to determine not just that one policy works, but also why it does so (Deaton and Cartwright 2018, 11–2). Any given experimental or non-experimental study may be limited in its capacity to unravel all the necessary conditions that must obtain for a policy to achieve a certain outcome. Looking across different types of evidence, however, allows us to think beyond the results of any given study to determine what must be tweaked to improve positive results (e.g., ensure that a policy or treatment will work the vast majority of time for the vast majority of individuals, not just on average) or correct negative ones. Systematic reviews, such as those compiled by the Cochrane Collaboration (for medical interventions) or the Campbell Collaboration (for policy interventions) reveal the extent of consensus across the available studies, as well as exposing gaps in the research.

Our aim in this section was modest. We argued that well-conducted randomized trials yield considerable epistemic value; and in light of that, they are a useful tool in the choice of policy. We now move on to discussing the politico-democratic value of

13 Blinding ensures participants will not know whether they were allocated to the treatment or control groups.

randomized trials. We argue that in comparison to other types of evidence, randomized trials most readily jointly advance a trio of key democratic principles: non-arbitrariness, revisability and public justifiability.

4 The Political Theory of Randomized Experiments

Leaving aside any design-related epistemic advantages or disadvantages, let us now focus more closely on the relationship between randomized policy experiments and democracy.

4.1 Non-arbitrariness: Randomized Trials Prevent the Abuse of Political Power

The first democratic principle that well-conducted randomized trials advance is non-arbitrariness of political power. Democracy and randomized trials both promote non-arbitrary political rule, albeit differently. Democratic institutions advance non-arbitrariness by ensuring that those exercising political power do so on the authority of, and in ways that are accountable to, the rest of the community. In democracies, political power in general is 'non-arbitrary' in that sense. But there is a second, different, sense in which any given discrete exercise of political power—any given law or policy—can be said to be non-arbitrary, regardless of whether it has been explicitly authorized by democratic processes or not.[14] That second way a particular policy can be non-arbitrary is by being objectively justified.

Insisting on the non-arbitrariness of discrete laws or policies addresses two problematic aspects of political power. First, political power is coercive power that can impose limitations on individual autonomy (Rawls 2005, 68). Thus the benefits of solving particular social problems through regulation must outweigh the cost to individual autonomy that such regulation would impose on individuals. The less arbitrary political power is, the more justifiable any restrictions to personal autonomy will be.[15] Second, exercised in an arbitrary way, political power can unfairly discriminate

14 Indeed, few laws or policies enjoy specific democratic authorization for the simple reason that during elections citizens vote on bundles of policies (as contained in the parties' electoral programs). Occasionally, a law or policy might receive specific democratic authorization through a referendum.
15 Political power may be arbitrary not only when it is exercised on a whim, for *no* reason. It may be arbitrary also when exercised for the *wrong* reasons. Any exercise of power that is not within the bounds of its authority—that is, that it is not responding to the reasons why the power was vested in that agent in the first place—may be deemed 'arbitrary'. Goodin 2004, 300–2.

between individuals or groups of citizens. Policies that advantage or disadvantage certain groups or individuals can be considered 'unfair treatment' if they are adopted in an arbitrary manner. The values of autonomy and fairness thus both require that political power be exercised in an objectively non-arbitrary way.

What does that mean when it comes to policy? Governmental agencies in the United States, for example, must act in accordance with the Administrative Procedure Act. Under that law, agency decisions can be subjected to an 'arbitrary and capricious' review and struck down by courts. Their regulation can be deemed an 'abuse of discretion' (Administrative Procedure Act, 5 U S C Code § 706, 2A) if, for example, it is not justified by reference to costs and benefits.[16] While it matters that projected benefits outweigh anticipated costs, there is a more basic sense in which policy can be arbitrary. This is if the policy in question does not provide a solution at all to the problem it is supposed to address, and hence is unable to secure the presumed benefits. A regulation that does not solve any social problem can be deemed an arbitrary exercise of political power. Needless to say, ineffective policies that impose costs and risks of harm will be especially problematic democratically, by this standard.

While under this Act, ineffective policies are not currently within the scope of legal review and thus unlikely to be deemed arbitrary by courts, they are nonetheless ticking bombs from a liberal democratic point of view. It is not just that such arbitrary policies are democratically deficient in gratuitously restricting autonomy and potentially engaging in unfair treatment. They also have the more insidious effect of eroding citizens' trust in their institutions and liberal democracy overall. Citizens are also less likely to comply with legal bans or mandates if they have reasonable grounds for doubting that such regulations will make a difference. Problem-solving is thus crucial to a government's capacity to exercise power non-arbitrarily. Policies that are arbitrary or ineffective may be adopted out of error, incompetence or political calculations.[17] While we should not neglect political considerations, quite often ineffective policy will stem from epistemic deficits (King and Crewe 2013).

There are however ways to enhance agencies' problem-solving capacity. While policymaking is increasingly evidence-based, testing policy through randomized trials offers additional benefits, as set out in Section 3. It can provide policymakers with evidence that any given existing policy may not in fact be working, as

16 *Michigan v. EPA*, 135 S. Ct. at 2711; Bus. Roundtable v. Sec. Exch. Comm'*n*, 647 F. 3d 1144, 1148–49 (D.C. Cir. 2011). See Sunstein 2017 for a discussion.

17 Political leaders' *a priori* judgements regarding efficiency might be unduly influenced by the interests they have in maintaining or changing any particular policy. Their judgements might also be clouded by cognitive biases such as loss aversion, whereby they become invested in existing policies and thus eager to maintain the status quo for no good reason.

previously claimed. Additionally, it can help policymakers adopt new, better solutions to existing problems, including when little prior evidence is available and there is need to somehow prove concept without imposing unjustified risks on the entire population. Unlike quasi-experimental evaluations, the researcher conducting a randomized trial can design it to test specific theories. This makes it possible for a randomized trial to shed light on the causal pathways through which a particular program works.

In recent years, randomized trials have suggested that certain programs are less effective than their proponents claim, often on the basis of low-quality evaluations, such as before–after comparisons. Among the social programs that have been shown by randomized trials to be ineffective are: Scared Straight (a program that exposes juvenile delinquents to a brief spell in prison); the 21st Century Community Learning Center initiative (a program of after-school activities for at-risk children); and abstinence-only programs aimed at reducing teen pregnancy. These programs were supported by quasi-experimental evaluations, but randomized evaluations suggested that they were not achieving their stated goals (DiCenso et al. 2002; James-Burdumy et al. 2004; James-Burdumy, Dynarski, and Deke 2008; Petrosino and Turpin-Petrosino 2002; Underhill, Montgomery, and Operario 2007).

By setting a more rigorous benchmark, randomized evaluations often fail to show statistically significant evidence of positive policy effects. Among the randomized trials commissioned by the US What Works Clearinghouse, just one in 10 produced any positive effects (Leigh 2018, 77). A randomized trial of online diversity training concluded that it is not a panacea for remedying bias in the workplace (Chang et al. 2019). In contrast to prior evidence from natural experiments, six randomized evaluations of microcredit programs in developing countries found that it is modestly positive, but not transformative (Banerjee, Karlan, and Zinman 2015). A randomized evaluation of an employer-sponsored wellness program concluded that it had no positive impacts on gym visits, medical spending or employee retention (Jones, Molitor, and Reif 2019). Usefully, the authors of the wellness program evaluation also showed that a non-randomized evaluation would have produced quite different results: wrongly implying that the program doubled the odds of participants visiting the gym, while halving the chances that they left the organization. This illustrates the value of randomized trials: in helping us reconsider our assessments of particular policies they can help us avoid arbitrary exercises of political power.

Well-conducted randomized trials in medicine, along with evidence on safety and efficacy, serve a similar purpose. Only one in 10 of the drugs that look promising in laboratory tests ends up passing through Phase I, and then Phase II and Phase III randomized trials and getting approval from the US Food and Drug Administration. This matters for public policy because, across the OECD, 71 percent of health costs are borne by taxpayers (even in the United States, the share is 50

percent) (OECD 2019, 159). Randomized trials help to ensure that public funds are not misspent on pharmaceuticals that would have harmed or delivered no real benefit to patients. Randomized trials in other areas of medicine have also helped to save lives and reduce wasteful expenditure.[18] Given that a substantial share of surgical costs is borne by taxpayers, surgical randomized trials help ensure that the government is not funding ineffective interventions. The same argument could extend to a wide range of government policies and programs whose costs are similarly borne by the citizen community.

In response to our argument that randomized trials help to prevent the arbitrary use of political power, some reply that randomized trials themselves are fundamentally arbitrary. After all, they employ arbitrariness in their design by randomly assigning participants to treatment and control groups. But assignment is not based on a whim or caprice. The structure of a randomized trial is grounded in scientific inquiry, and is justifiable to participants and the broader community. Randomized trials are not conducted for arbitrary purposes (you need to have some reason to think that the treatment is going to work in order to justify enrolling people in a trial) nor do they use coercive state power (participation in trials that involve a significant burden on the participant typically requires informed consent). Finally, while a small subset of the wider population may be randomly ('arbitrarily') selected to participate in the trial, if the trial succeeds and policy is changed accordingly, the general population will receive just the same benefits as the trial participants.

In a nutshell, randomized trials help to ensure that policies will be effective, and that they do not harm participants. In doing so, together with other good quality evidence, they uphold the principle of non-arbitrariness, according to which political power should be exercised in ways that do not gratuitously frustrate personal autonomy and interpersonal fairness.

4.2 Revisability: Democracy Provides Fertile Ground for Randomized Trials

There is a second way in which democracy and randomized trials are almost inextricably interrelated. In challenging the status quo, randomized trials serve

18 Lung volume reduction surgery, for example, was once used to treat patients with severe emphysema, until a randomized trial showed that it significantly increased the risk of death. Other examples of inefficient (and often harmful) medical procedures that have been phased out thanks to randomized trials include an extracranial-to-intracranial bypass after a minor stroke and laparoscopic surgery to 'unpick' the adhesions in patients whose bowels were caught up in scar tissue. For a survey, see Wartolowska et al. 2014.

democracy's defining commitment to revisability. That same commitment makes democracy, par excellence, the type of regime that can in practice rely on randomized trials for policy purposes, without undermining its stability or indeed survival. To some extent, all good science serves democracy's commitment to revisability. Yet, as we will argue in the next section, randomized trials are perhaps the most open and transparent way in which a government can test its policies. Due to their conspicuous procedural impartiality and simple design, randomized trials are the most transparent way in which a regime can evaluate itself. Randomized trials can also be ill-conducted. But those failures will tend to be more apparent and harder to conceal than the results of policy evaluations that use techniques that are complex and readily adulterated.

It is a telling fact that randomized trials were virtually non-existent under Soviet communism. McKee (2007) reports:

> There were two reasons why randomized controlled trials did not thrive under communism. One was the politicization of Soviet science, in which scientific uncertainty was not accepted and where methodological limitations were overlooked as long as research complied with the prevailing communist ideology. But the second was that it served the needs of the system to prevent established solutions from being questioned. If they were, and were found wanting, what would be left? (271)

The open-mindedness underpinning randomized trials conflicted directly with the ethos of the Soviet communist regimes. Randomized trials would have more easily introduced an element of doubt that had the potential to challenge their authority. Authoritarians rarely welcome the prospect of seeing their cherished ideas refuted.

In classifications of regimes, countries can be arrayed along a spectrum from full autocracy to full democracy (see e.g. Gurr 1974). Some countries hold regular elections, yet the same party always seems to win. It turns out that even on this dimension of democracy, it is possible to see differences in the extent to which nations conduct randomized trials. The less politically competitive the election (i.e. the larger the vote margin), the fewer randomized trials that a country conducts (Corduneanu-Huci, Dorsch, and Maarek 2021). The pattern holds even within countries—as nations become less electorally competitive, they conduct fewer randomized experiments.

When leaders face no serious opposition, there is less incentive to engage in evidence-based policymaking in order to persuade voters or convince potential political allies. Leaders with autocratic tendencies are more likely to ignore randomized trial results—witness the way in which Presidents Donald Trump and Jair Bolsonaro continued to tout hydroxychloroquine as a treatment for COVID-19, even after

randomized trials showed it to be ineffective (Cavalcanti et al. 2020; Skipper et al. 2020).[19] When leaders feel that they need to justify their policies to the media, civil society and the electorate, they are more likely to rely upon evidence from randomized trials.

It is no accident that less democratic regimes tend to conduct fewer randomized trials. Policy experiments are more likely to emerge from an environment that is open to change. In testing policy, decision-makers must be ready to accept that the policy in question might be wrong. There is no point experimenting with alternatives if there is no possibility of changing course. Policy experiments are thus better suited to a political regime that embraces revisability.[20]

Liberal democracy fits this profile. It is founded upon ideas of openness to change. Liberalism (both as a personal philosophy and a political ideology) endorses revisability, rejecting dogma and blind faith. It approaches questions with a willingness to change one's mind. 'The spirit of liberty', as Judge Learned Hand once put it, 'is the spirit that is not too sure it is right' (Hand 1944). And democracy is by definition open to change. Citizens can always change their minds; parties can always be voted out of office. As important as respecting the will of the people is respecting the fact that this will can change. This is why democracies hold periodic elections. And this is why peaceful transitions of power, not just free and fair elections, are truly markers of democracy (Popper 1980; Przeworski 2015).

A state willing to experiment with policy is one that is open to changing it. A government that is willing to test its preferred policy through a variety of methods is one that is ready to accept that it can be wrong.[21] It is also one that is ready to accept the electoral consequence of such mistaken choice: the risk of losing political power. Just as the *randomista* must be prepared to accept that upon being tested a treatment may turn out to be ineffective, a government engaging in experimental policy testing must be ready to accept that, in the face of such evidence, it can be proven to voters to have been wrong. We are talking of a governing party that embraces two types of

19 Trump's enthusiasm was partly due to the results of a small non-randomized trial in France that found that hydroxychloroquine was effective in treating patients with COVID-19 (Gautret et al. 2020).

20 Or, in Popper's terms, that adopts the 'piecemeal social engineering approach' permitting 'repeated experiments and continuous adjustment' (1980, 163). In public policy this is known as the doctrine of incrementalism (Braybrooke and Lindblom 1963; Wildavsky 1979; see Goodin 1982, ch. 2 for a discussion).

21 Policy experimentation characteristic to piecemeal social engineering 'might lead to the happy situation where politicians begin to look out for their own mistakes instead of trying to explain them away and to prove that they have always been right. This—and not utopian planning or historical prophecy—would mean the introduction of scientific methods into politics, since the whole secret of scientific method is a readiness to learn from mistakes' (Popper 1980, 163).

revisability. First, it must be open to revising its preferred policies.[22] Second, it must be open to the possibility of political change—of political revisability that is intrinsic to democracy—that voters may change their minds about which party is best fit to rule.

Public policies, embodying commonly held beliefs, can be tested through randomized trials. Randomized trials are thus also an important step toward changing policies if they are found wanting after weighing all available evidence.

Take for example the question of whether bed nets to prevent malaria in developing countries should be distributed at a low price or at zero cost. Some experts had argued that free bed nets often ended up being misused as fish nets or wedding veils, and that recipients would be more likely to sleep under a bed net if they were required to pay a low price (e.g., Easterly 2006, 12). Accordingly, in the period 2000 to 2005, the World Health Organization focused on distributing subsidized bed nets, costing USD 2–3 each (Sachs 2009). Then a series of randomized trials demonstrated that recipients were equally likely to sleep under free bed nets as low-priced bed nets. Moreover, take-up was considerably higher when the nets were free (even at USD 0.60, demand fell by two thirds) (Cohen and Dupas 2010; Dupas 2009; Dupas 2014). As a result of the randomized trials, the World Health Organization switched its focus to favor distributing free bed nets (Sachs 2014). The change has averted many thousands of malaria deaths.

In science, Kuhn (1962) discussed how fields move from periods of normal science (in which widely shared assumptions govern research) to sudden scientific revolutions (in which orthodoxies are overthrown). Public policy should also be capable of making major shifts when necessary. Something is lost when a community is incapable of changing course in light of new high-quality evidence. Neglecting well-conducted randomized trials has the same costs that silencing speech has. As Mill put it: 'If the opinion is right, they [citizens] are deprived of the opportunity of exchanging error for truth; if wrong, they lose, what is almost as great a benefit, the clearer perception and livelier impression of truth produced by its collision with error' (Mill 1859, 33). Similarly, a world without randomized trials would not only be one in which it is harder to prove the efficacy of certain policies, it is also one in which we cannot disprove 'common wisdom' that is in fact false.

Policy experimentation can occur only insofar as a political regime and its agents are open to change. In endorsing revisability, liberal democracy makes such experimentation possible. Revisability is thus the first democratic principle to which randomized trials cater. But any given randomized trial advances revisability only partially. Its results, if conclusive, may trigger policy change. Yet in a deeper sense, revisability requires that we also be open to evidence-based reversal of this new policy

22 In other words, they should advocate the importance of the problem, rather than the importance of the solution—the latter should be open to change (Campbell 1969, 427).

at a later date. Policymakers should be cautious about seeing any given randomized trial (or, more generally, any given piece of evidence) as the 'final word' on a topic. If so, randomized trials would be serving the cause of revisability only imperfectly. At best, they would promote what we might call a 'closed' type of revisability.[23]

'Open' perpetual revisability of policy can occur in several ways. Monitoring can be done through the collection and analysis of both experimental and non-experimental data by governmental agencies and researchers. Subsequent studies —both experimental and non-experimental—may reanalyze the results and uncover fresh findings (for example, discovering that a housing mobility policy had a larger impact on toddlers than teenagers) (Chetty, Hendren, and Katz 2016). Replication studies might, for example, find smaller policy impacts than the original randomized trial, thus casting doubt on whether the program should indeed be scaled up.

4.3 Public Justification: Randomized Trials can Help Win Public Support

According to the principle of non-arbitrariness, policy is legitimate in virtue of what could justify it from an objective perspective. Yet its democratic legitimacy will also hinge on what could justify it from citizens' subjective perspective.[24] Think, for example, of abstinence-only sex education programs or a policy of distributing free bed nets. Citizens may have very strong prior beliefs on those topics. To shift those views, it is helpful for policymakers to draw on evidence the public will easily comprehend and find compelling. Contrary to what many citizens might have expected, evidence from randomized trials showed abstinence-only sex education to be ineffective at preventing unwanted teen pregnancy and distributing free bed nets to be effective in saving lives.

According to the principle of public justification, coercive power always stands in need of public justification (Feinberg 1987, 9; Rawls 2005; Scanlon 1998). This means that policies should respond to reasons that citizens could reasonably accept (Rawls 2005; Scanlon 1998). The objective effectiveness of policy is one such reason. We have already discussed that in relation to non-arbitrariness, where we pointed out that any kind of quality evidence can advance this goal. But randomized trials are particularly well-placed to help expand the list of public justifications beyond

23 Our terminology is inspired by Sen's (2002) distinction between open (universal) impartiality and closed impartiality.

24 On this account, it matters less that policymakers were justified in adopting a policy. What matters is that citizens themselves could find reasons for endorsing that policy.

such reason, thereby ensuring that policies can be endorsed by larger swathes of the citizenry. This in turn explains why democracies should particularly want to include high-quality randomized trials in their policymaking armory. While not all policies can be tested through such trials, we believe that many policies can benefit from this sort of testing.

4.3.1 'People like Us' and the Justifiability of Policy

Procedural fairness is one important aspect making lotteries appealing. Before the drawing of the random sample begins, every potential participant has an equal chance of being selected. This is why governments sometimes use lotteries to distribute scarce goods (such as housing vouchers and charter school places) or to impose unwelcome burdens (such as the Vietnam draft).[25] But there is a second aspect contributing to lotteries' appeal: random selection can ensure a descriptive identity, in a statistical sense, between those who win the lottery and those who do not. The winners can be seen as proxies for the rest of the community.

Lottocracies choose rulers randomly by lot from among the wider citizenry, making it more likely that rulers will be similar to the average citizen (Amar 1984; Burnheim 1985; Callenbach and Phillips 1985; for a discussion, see Stone 2011, 132–41). Deliberative Polls likewise select deliberators randomly from among the citizenry, while also using stratified sampling to ensure all demographic groups are properly represented. The aim, in both cases, is to ensure that the deliberative group will be a microcosm or mirror of the entire community (Fishkin 2009, 24–9, 82). The hope is that this close identity will in turn enhance the public justifiability of collective decisions reached by randomly selected decision-making bodies.

In providing public reasons based on the similarity between trial participants and the wider public, large randomized trials enjoy the same justificatory advantages as sortition and Deliberative Polling. By randomly selecting and assigning participating citizens to treatment and control groups, randomized trials can test policy on a microcosm of the entire community. The fact that the policy was tested on similarly situated citizens may constitute a more compelling reason for the rest of the citizenry to have faith that the estimated effects would generalize to them.[26]

---

25 See Stone 2007 for a discussion of lotteries and allocative justice.

26 While the average effects estimated from a policy evaluation provide an informative prior, there will inevitably be heterogeneity across the population (as the automobile advertisements of the 1970s famously noted 'your mileage may vary'). But many trials also report results for sub-groups (broken down, for example, by age or gender), and even a randomized trial that reports only a single estimate provides valuable guidance, provided that it is based on a large, representative sample.

Take for example, the 2008 Oregon Health Insurance Experiment, which involved the state government extending public health insurance to an additional 10,000 low-income families (Finkelstein et al. 2012). Facing a situation in which there were nearly nine families for every available place in the program, the state allocated the spots through a public lottery. The trial showed that insurance led to people reporting they felt physically healthy two more weeks a year and mentally healthy three additional weeks a year. The lottery was designed to ensure that the allocation would be perceived by the rest of the non-participating families as a 'fair' one. But it also had the further effects of providing other Oregon families with a compelling reason for supporting public health insurance. Since the experiment had been conducted on other Oregon families just like them, it was easy for participants to see that the extension of public health insurance would be beneficial to them.

Finding direct evidence that citizens are more likely to endorse randomly trialed policies due to a perceived identity between them and trial participants is difficult. Yet, political theorists have long argued a similar point about Deliberative Polls and mini-publics: that the statistical similarity between the wider public and randomly selected deliberators will increase the public legitimacy of the subsequent collective decisions. This is one of the long-standing arguments made in favor of randomized deliberative participatory decision-making (Fishkin 2009, 35, 151–2, 131; Warren 2008; for a discussion, see also Lafont 2019). And recent evidence does indeed suggest that citizens are more like to accept a decision the closer the identity they perceive between themselves and the deliberators.[27]

Mutatis mutandis, policies adopted on the basis of large randomized trials using random selection (similarly to decisions made by lottocratic rulers or adopted through Deliberative Polling), will arguably enjoy the same justificatory advantage, being able to provide citizens extra, special reasons for accepting policies, which are grounded in the descriptive identity between themselves and trial participants.

4.3.2 Democratic Scrutability of Evidence

It is also important to notice that, in terms of supplying public justification, randomized trials will generally be superior to other types of scientific evidence. This is because, epistemically, they are accessible to lay citizens in ways other evidence is not.[28] The methodology of randomized trials is not as mysterious for the average

27 Pow, van Dijk, and Marien 2020, 50–1; for related discussions, see also Bedock and Pilet 2021; Garry et al. 2022; Goldberg and Bächtiger 2021; Werner and Marien 2022.

28 Burtless (1995, 69–70, 72) advances a similar argument about senior officials and political incumbents. Because of their simplicity, randomized trials results are more accessible to and thus also more convincing to bureaucrats and politicians. The latter are also more likely to act on the

citizen as propensity score matching, regression discontinuity or instrumental variables. Citizens may not easily understand concepts such as selection effects, but they can understand differences in outcomes for people randomly assigned to control and treatment branches of a trial. The general idea behind a randomized trial is easier to understand by the average citizen than a study based on quasi-experimental techniques such as differences-in-differences, regression discontinuity or instrumental variables.[29]

Moreover, because the design of randomized trials does not rely on many prior assumptions, *ceteris paribus*, such evidence can be especially valuable socially and politically when citizens are polarized in their beliefs and hence do not share the same priors.[30] When scientific evidence is—and more importantly, when it is perceived as being—epistemically accessible and neutral, it can supply more robust public justifications for a wider set of the community. Such evidence can thus achieve democratic, not just expert, scrutability; enhancing thereby the democratic legitimacy of policy. In this sense, the case for randomized trials is akin to one of the earliest notions of science: that it represents a form of democratic knowledge, which is accessible to everyone, regardless of education or social background (Fuller 1997). In this sense, randomized trials hark back to an older tradition in liberal democracy—a form of accountability in which outcomes are observable in the public square. Although the complexity of modern public policy often renders this kind of accountability impossible, randomized trials may satisfy this test, providing the public with readily observable evidence of what works and what does not (Ezrahi 1990, 286; Pearce and Raman 2014).

Advocates of some policy reform often provide evidence of various sorts to the public. Citizens may, however, have simply taken on board the conclusions of the evidence presented, without understanding very well how those conclusions were reached. They may have simply trusted those conclusions, since the way those conclusions were reached is not transparent to them. That situation is better than one where no such evidence is provided, but it does require some epistemic 'leap of faith' on the part of the citizens. Even where such a leap of faith is epistemically

---

results since no other complicated qualifications apply to them. This is arguably one reason why the findings of the Work-Welfare Experiments, commissioned by the Manpower Demonstration Research Corporation, had an important impact on the design and implementation of the 1988 Family Support Act (Burtless 1995, 70). Deaton and Cartwright (2018, 10) similarly acknowledge that randomized trials can most easily be used as 'dispute-reconciliation mechanisms to resolve political conflicts'.

29 In extremely rare cases, researchers have used statistical adjustments to correct the results of randomized trials. For example, Bijwaard and Ridder (2005) propose an econometric correction for the case of selective compliance by those in the treatment group.

30 This advantage of randomized trials is even acknowledged by Deaton and Cartwright (2018, 10).

warranted and likely to enhance citizens' competence, something is lost nonetheless from the standpoint of public justification. Citizens' appreciation of a study will be greater if they are able to critically engage and understand how it worked. The more epistemically transparent and accessible the provided evidence is, the greater its capacity to offer public justification to citizens for any given policy. In a different context, compare the statements: '99 percent of scientists believe the climate is warming' with 'a twelve-year-old child has experienced nine of the 10 warmest years on record'. The former is persuasive only to those who respect the views of scientists, while the latter relies only on the listener accepting the accuracy of the temperature record. Compared with more complex evaluation tools, randomized trials may be perceived as closer to 'argument by fact' than to 'argument by authority'.

In virtue of this, randomized trials are particularly capable of publicly justifying policy. Their simple setup and general methodology allow them to more actively engage citizens epistemically. Citizens are not just asked to trust researchers' conclusions—something that risks disempowering citizens' epistemic agency, treating them as what we may call 'epistemic patients'. Instead of circumventing citizens' own independent thinking, randomized trials speak directly to it. This is something that other types of research relying on complex statistical analysis or modelling have difficulty doing. There is thus another special epistemic sense in which randomized trials treat their audience fairly. This is by how they engage their audience's epistemic agency—the audience's capacity to think about and understand scientific evidence. Randomized trials can thus achieve democratic scrutability—the capacity of being understood by a large portion of the community.

5 A Design Proposal: The Role of Citizens and Experts

Having discussed the democratic advantages of randomized trials, we are left to wonder: can the process by which randomized trials are initiated be made more democratic? One reason for doing this is that large randomized trials can be costly. Providing some democratic input into the choice regarding which such trials should be undertaken would be good, especially where such trials are conducted by governmental agencies and thus their cost is borne by citizens.

Invariably, many ideas for randomized trials will emerge from public service agencies (who seek to evaluate their programs), or from the research community (who seek to better understand human behavior). Just as the pipeline of medical randomized trials will largely be determined by drug companies and research at

the frontier of knowledge, so too randomized trials of public policy proposals will be shaped to a large extent by policymakers and experts. There is thus a risk that such policy evaluation will just be epistocracy in new clothes: more power to experts will mean less for citizens. But is there also an opportunity for citizens to provide input into which programs should be evaluated through randomized trials?

One possible answer would be to use randomized policy evaluation as an opportunity for giving power to citizens. Similarly to how deliberative mini-publics like the Oregon Citizens' Initiative Review have been used to choose which propositions should be put to a vote in an initiative or referendum, such randomly selected deliberative mini-publics could be used to decide at least some of the policies to be subjected to this kind of thorough publicly funded random testing. This process might identify established programs whose effectiveness is widely accepted among experts, but doubted among citizens. Or it might involve novel policies that challenge accepted wisdom. The process could also be an iterative one, with the results from randomized trials informing further deliberative democracy processes.

Such a process might be most effective if conducted at arm's length from government—under the auspices of, for example, a foundation, think tank or university. Although this would limit the extent to which randomized trials could be developed in collaboration with government, it would have the advantage of providing a greater sense of public justification, and potentially greater citizen buy-in to the process.

6 Conclusion

Policymaking already entails a good deal of ad-hoc experimentation. As parties rotate in office, they subject the entire population to different policies. Pilot programs are ubiquitous, but all too frequently it is difficult to discern the counterfactual: what would have happened to those in the program if they had not received the treatment? One way to think about well-conducted randomized trials is that they amount to a pilot program, but one with a credible control group. Evidence syntheses, which draw together the available evidence (randomized and quasi-experimental), can help policymakers make better decisions.

Empirically, randomized trials turn out to be more common in more democratic countries. In nations where power changes hands regularly at the ballot box, there tend to be more randomized trials. We argue that this is because liberal democracies founded upon the idea of revisability and change are especially suited for randomized policy testing, which provides citizens and political leaders with evidence about how to form and reform policy.

Randomized trials offer unique and non-unique democratic benefits. Together with other types of evidence, randomized policy trials can also ensure that coercive political power will not be exercised in an arbitrary way that frustrates personal autonomy gratuitously. But randomized trials are uniquely well-placed to enhance the public justifiability of policy. Their democratic scrutability and their reliance on citizens' participation may provide the larger public with a wider set of reasons for embracing a policy.

Acknowledgements: We thank Bob Goodin, Jeevan Haikerwal, John Langmore and participants at a Melbourne University symposium on 'Social Liberalism in Contested Space' for valuable comments on earlier drafts.
Research funding: Ana Tanasoca acknowledges financial support from a Macquarie University Research Fellowship.

References

Administrative Procedure Act, 5 U.S.C § 500 et seq. Also available at: https://www.law.cornell.edu/uscode/text/5/part-I.

Amar, A. R. 1984. "Choosing Representatives by Lottery Voting." *The Yale Law Journal* 93 (7): 1283–308.

American Economic Association (AEA). 2021. *RCT Registry: Registry for Randomized Controlled Trials*. Also available at: https://www.aeaweb.org/journals/policies/rct-registry.

Backmann, M. 2017. "What's in a Gold Standard? In Defense of Randomized Controlled Trials." *Medical Health Care and Philosophy* 20: 513–23.

Banerjee, A., D. Karlan, and J. Zinman. 2015. "Six Randomized Evaluations of Microcredit: Introduction and Further Steps." *American Economic Journal: Applied Economics* 7 (1): 1–21.

Bedock, C., and J.-B. Pilet. 2021. "Who Supports Citizens Selected by Lot to Be Main Policymakers? A Study of French Citizens." *Government and Opposition* 56: 485–504.

Bijwaard, G. E., and G. Ridder. 2005. "Correcting for Selective Compliance in a Re-employment Bonus Experiment." *Journal of Econometrics* 125 (1–2): 77–111.

Bloom, H. S., C. Michalopoulos, C. J. Hill, and Y. Lei. 2002. Can Nonexperimental Comparison Group Methods Match the Findings from a Random Assignment Evaluation of Mandatory Welfare-To-Work Programs? In *MDRC Working Papers on Research Methodology*.

Braybrooke, D., and C. E. Lindblom. 1963. *A Strategy for Decision*. New York: Free Press.

Burnheim, J. 1985. *Is Democracy Possible? The Alternative to Electoral Democracy*. Cambridge: Polity Press.

Burtless, G. 1995. "The Case for Randomized Field Trials in Economic and Policy Research." *The Journal of Economic Perspectives* 9 (2): 63–84.

Callenbach, E., and M. Phillips. 1985. *A Citizen Legislature*. Berkeley: Banyan Tree books.

Campbell, D. T. 1969. "Reforms as Experiments." *American Psychologist* 24 (4): 409–29.

Campbell, D. T. 1998. "The Experimenting Society." In *The Experimenting Society: Essays in Honor of Donald, T. Campbell, Policy Studies Review Annual*, Vol. 11, edited by W. Dunn, 35–68. New Brunswick: Transaction Publishers.

Cavalcanti, A. B., F. G. Zampieri, R. G. Rosa, L. C. Azevedo, V. C. Veiga, A. Avezum, L. P. Damiani, A. Marcadenti, L. Kawano-Dourado, T. Lisboa, D. L. and Junqueira. 2020. "Hydroxychloroquine with

or without Azithromycin in Mild-To-Moderate Covid-19." *New England Journal of Medicine* 383 (21): 2041–52.

Chang, E. H., K. L. Milkman, D. M. Gromet, R. W. Rebele, C. Massey, A. L. Duckworth, and A. M. Grant. 2019. "The Mixed Effects of Online Diversity Training." *Proceedings of the National Academy of Sciences* 116 (16): 7778–83.

Chetty, R., N. Hendren, and L. F. Katz. 2016. "The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment." *The American Economic Review* 106 (4): 855–902.

Cohen, J., and P. Dupas. 2010. "Free Distribution or Cost-Sharing? Evidence from a Randomized Malaria Prevention Experiment." *Quarterly Journal of Economics* 125: 1–45.

Corduneanu-Huci, C., M. T. Dorsch, and P. Maarek. 2021. "The Politics of Experimentation: Political Competition and Randomized Controlled Trials." *Journal of Comparative Economics* 49 (1): 1–21.

Deaton, A., and N. Cartwright. 2018. "Understanding and Misunderstanding Randomized Controlled Trials." *Social Science and Medicine* 210: 2–21.

Dewey, J. 2012. *The Public and its Problems*. University Park: Pennsylvania University Press.

DiCenso, A., G. Guyatt, A. Willan, and L. Griffith. 2002. "Interventions to Reduce Unintended Pregnancies Among Adolescents: Systematic Review of Randomized Controlled Trials." *British Medical Journal* 324 (7351): 1426.

Dupas, P. 2009. "What Matters (And What Does Not) in Households' Decision to Invest in Malaria Prevention?" *The American Economic Review: Papers and Proceedings* 99 (2): 224–30.

Dupas, P. 2014. "Short-run Subsidies and Long-Run Adoption of New Health Products: Evidence from a Field Experiment." *Econometrica* 82 (1): 197–228.

Easterly, W. 2006. *The White Man's Burden: Why the West's Efforts to Aid the Rest Have Done So Much Ill and So Little Good*. New York: Penguin.

Evidence Based Medicine Working Group. 1992. "Evidence-based Medicine: A New Approach to Teaching the Practice of Medicine." *The Journal of the American Medical Association* 268: 2420–5.

Ezrahi, Y. 1990. *The Descent of Icarus: Science and the Transformation of Contemporary Democracy*. Cambridge: Harvard University Press.

Fuller, S. 1997. *Science*. Buckingham: Open University Press.

Feinberg, J. 1987. *Harm to Others: The Moral Limits of the Criminal Law*. New York: Oxford University Press.

Fiorina, M. P. 1981. *Retrospective Voting in American National Elections*. New Haven: Yale University Press.

Finkelstein, A., S. Taubman, B. Wright, M. Bernstein, J. Gruber, J. P. Newhouse, H. Allen, and K. Baicker, and Oregon Health Study Group. 2012. "The Oregon Health Insurance Experiment: Evidence from the First Year." *Quarterly Journal of Economics* 127 (3): 1057–106.

Fishkin, J. S. 2009. *When the People Speak*. Oxford: Oxford University Press.

Garfinkel, I., C. F. Manski, and C. Michalopoulos. 1992. "Micro Experiments and Macro Effects." In *Evaluating Welfare and Training Programs*, edited by C. F. Manski, and I. Garfinkel, 253–76. Cambridge: Harvard University Press.

Garry, J., J. Pow, J. Coakley, D. Farrell, B. O'Leary, and J. Tilley. 2022. "The Perception of the Legitimacy of Citizens' Assemblies in Deeply Divided Places?" Evidence of Public and Elite Opinion from Consociational Northern Ireland." *Government and Opposition* 57 (3): 532–51.

Gautret, P., J. C. Lagier, P. Parola, L. Meddeb, M. Mailhe, B. Doudier, J. Courjon, V. Giordanengo, V. E. Vieira, H. T. Dupont, and S. Honoré. 2020. "Hydroxychloroquine and Azithromycin as a Treatment of COVID-19: Results of an Open-Label Non-randomized Clinical Trial." *International Journal of Antimicrobial Agents* 56 (1): 105949.

Gleason, P. M., A. M. Resch, and J. A. Berk. 2012. *Replicating Experimental Impact Estimates Using a Regression Discontinuity Approach*. National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. NCEE Reference Report 2012–4025.

Glewwe, P., and K. Muralidharan. 2016. "Improving Education Outcomes in Developing Countries—Evidence, Knowledge Gaps, and Policy Implications." In *Handbook of the Economics of Education*, Vol. 5, edited by E. Hanushek, S. Machin, and L. Woessman, 653–744. Amsterdam: North Holland.

Global Commission on Evidence to Address Societal Challenges. 2022. *The Evidence Commission report: A wake-up call and path forward for decisionmakers, evidence intermediaries, and impact-oriented evidence producers*. Hamilton: McMaster Health Forum.

Goldberg, S., and A. Bächtiger. 2021. "Catching the 'Deliberative Wave'? How (Disaffected) Citizens Assess Deliberative Citizen forums." Draft Manuscript.

Goodin, R. E. 1982. *Political Theory and Public Policy*. Chicago: University of Chicago Press.

Goodin, R. E. 1995. *Utilitarianism as a Public Philosophy*. Cambridge: Cambridge University Press.

Goodin, R. E. 2004. "Support with Strings: Workfare as an 'Impermissible Condition." *Journal of Applied Philosophy* 21 (4): 297–308.

Gurr, T. R. 1974. "Persistence and Change in Political Systems, 1800–1971." *American Political Science Review* 68 (4): 1482–504.

Hand, B. L. 1944. *The Spirit of Liberty. Speech Delivered in Central Park*. New York. (accessed May 21, 1944).

International Committee of Medical Journal Editors (ICMJE). 2021. *Clinical trials*. Also available at: www.icmje.org/recommendations/browse/publishing-and-editorial-issues/clinical-trial-registration.html.

James-Burdumy, S., M. Dynarski, and J. Deke. 2008. "After-school Program Effects on Behavior: Results from the 21st Century Community Learning Centers Program National Evaluation." *Economic Inquiry* 46 (1): 13–8.

James-Burdumy, S., M. Dynarski, M. Moore, L. Rosenberg, J. Deke, W. Mansfield, and E. Warner. 2004. *When Schools Stay Open Late: The National Evaluation of the 21st Century Community Learning Centers Program: Final Report*. US Department of Education, National Center for Education Evaluation and Regional Assistance.

Jones, D., D. Molitor, and J. Reif. 2019. "What Do Workplace Wellness Programs Do? Evidence from the Illinois Workplace Wellness Study." *Quarterly Journal of Economics* 134 (4): 1747–91.

Joyce, T. J., S. Crockett, D. A. Jaeger, O. Altindag, and S. D. O'Connell. 2014. Does Classroom Time Matter? A Randomized Field Experiment of Hybrid and Traditional Lecture Formats in Economics. In *NBER Working Paper 20006*. Cambridge: National Bureau of Economic Research.

Key, V. O. 1966. *The Responsible Electorate: Rationality in Presidential Voting, 1936-1960*. Cambridge: Harvard University Press.

King, A., and I. Crewe. 2013. *The Blunders of Our Governments*. London: Oneworld.

Kuhn, T. 1962. *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

Lafont, C. 2019. *Democracy without Shortcuts*. Oxford: Oxford University Press.

Landemore, H. 2013. "Deliberation, Cognitive Diversity, and Democratic Inclusiveness: An Epistemic Argument for the Random Selection of Representatives." *Synthese* 190 (7): 1209–31.

Leigh, A. 2018. *Randomistas: How Radical Researchers Are Changing Our World*. New Haven: Yale University Press.

McKee, M. 2007. "Cochrane on Communism: The Influence of Ideology on the Search for Evidence." *International Journal of Epidemiology* 36 (2): 269–73.

McKenzie, D., S. Stillman, and J. Gibson. 2010. "How Important Is Selection? Experimental vs. Non-experimental Measures of the Income Gains from Migration." *Journal of the European Economic Association* 8 (4): 913–45.

Meldrum, M. 1998. "'A Calculated Risk': The Salk Polio Vaccine Field Trials of 1954." *British Medical Journal* 317 (7167): 1233–6.
Mill, J. S. 1859. *On Liberty*. London: John W. Parker and Son.
Moffitt, R. 1992. "Evaluation Methods for Program Entry Effects." In *Evaluating Welfare and Training Programs*, edited by C. F. Mansky, and I. Garfinkel, 231–52. Cambridge: Harvard University Press.
Ober, J. 2017. *Demopolis: Democracy Before Liberalism in Theory and Practice*. Cambridge: Cambridge University Press.
OECD. 2019. *Health at a Glance 2019: OECD Indicators*. Paris: OECD Publishing.
Omotilewa, O., and J. Ricker-Gilbert. 2019. "Impact Evaluation of Large-Scale Extension Activities on Agricultural Technology Adoption: Experimental vs. Non-experimental Estimates." In *6th International Conference of African Association of Agricultural Economists (AAAE)*, Abuja, Nigeria, September 23–26, 2019.
Pearce, W., and S. Raman. 2014. "The New Randomised Controlled Trials (RCT) Movement in Public Policy: Challenges of Epistemic Governance." *Policy Sciences* 47: 387–402.
Petrosino, A., and C. Turpin-Petrosino. 2002. "Scared Straight' and Other Juvenile Awareness Programs for Preventing Juvenile Delinquency." (Updated C2 Review). In *Reviews of Intervention and Policy Evaluations*, (C2-RIPE). Philadelphia: Campbell Collaboration.
Popper, K. R. 1980. *The Open Society and its Enemies*, Vol. I. London: Routledge & Kegan Paul.
Pow, J., L. van Dijk, and S. Marien. 2020. "It's Not Just the Taking Part that Counts: 'Like Me' Perceptions Connect the Wider Public to Mini-Publics." *Journal of Deliberative Democracy* 16 (2): 43–55.
Przeworski, A.. 2015. "Acquiring the Habit of Changing Governments through Elections." *Comparative Political Studies* 48: 101–29.
Rawls, J. 2005. *Political Liberalism*. New York: Columbia University Press.
Rivlin, A. M. 1971. *Systematic Thinking for Social Action*. Washington: Brookings Institution.
Sachs, J. 2009. *Good News on Malaria Control*. Scientific American.
Sachs, J. 2014. "The Case for Aid." *Foreign Policy* (21 January).
Sauermann, J., and A. Stenberg. 2020. "Assessing Selection Bias in Non-experimental Estimates of the Returns to Workplace Training." In IZA Discussion Paper 13789. Bonn: IZA.
Scanlon, T. M. 1998. *What We Owe to Each Other*: Harvard University Press.
Schmeck, H. M. 1995. *Dr. Jonas Salk Whose Vaccine Turned Tide on Polio, Dies at 80*, 24 June, 1. New York Times.
Schumpeter, J. A. 1950. *Capitalism, Socialism and Democracy*, 3rd ed. New York: Harper & Row.
Sen, A. 2002. "Open and Closed Impartiality." *Journal of Philosophy* 99 (9): 445–69.
Skipper, C. P., K. A. Pastick, N. W. Engen, A. S. Bangdiwala, M. Abassi, S. M. Lofgren, D. A. Williams, E. C. Okafor, M. F. Pullen, M. R. Nicol, and A. A. Nascene. 2020. "Hydroxychloroquine in Nonhospitalized Adults with Early COVID-19: A Randomized Trial." *Annals of Internal Medicine* 173 (8): 623–31.
St.Clair, T., T. D. Cook, and K. Hallberg. 2014. "Examining the Internal Validity and Statistical Precision of the Comparative Interrupted Time Series Design by Comparison with a Randomized Experiment." *American Journal of Evaluation* 35 (3): 311–27.
Stone, P. 2007. "Why Lotteries Are Just." *The Journal of Political Philosophy* 15 (3): 276–95.
Stone, P. 2011. *The Luck of the Draw: The Role of Lotteries in Decision Making*. Oxford: Oxford University Press.
Sunstein, C. R. 2017. "Cost-benefit Analysis and Arbitrariness Review." *Harvard Environmental Law Review* 41 (1): 1–41.
Taleb, N. 2007. *The Black Swan: The Impact of the Highly Improbable*. New York: Random House.
Thaler, R. H., and C. R. Sunstein. 2009. *Nudge*. New York: Penguin.

Underhill, K., P. Montgomery, and D. Operario. 2007. "Sexual Abstinence Only Programmes to Prevent HIV Infection in High Income Countries: Systematic Review." *British Medical Journal* 335 (7613): 248–52.

Vinkers, C. H., H. J. Lamberink, J. K. Tijdink, P. Heus, L. Bouter, P. Glasziou, D. Moher, J. A. Damen, L. Hooft, and W. M. Otte. 2021. "The Methodological Quality of 176,620 Randomized Controlled Trials Published between 1966 and 2018 Reveals a Positive Trend but Also an Urgent Need for Improvement." *PLoS Biology* 19 (4): e3001162.

Warren, M. E. 2008. "Citizen Representatives." In *Designing Deliberative Democracy: The British Columbia Citizens' Assembly*. 50–69, edited by M. E. Warren, and H. Pearse. Cambridge: Cambridge University Press.

Wartolowska, K., A. Judge, S. Hopewell, G. S. Collins, B. J. Dean, I. Rombach, D. Brindley, J. Savulescu, D. J. Beard, and A. J. Carr. 2014. "Use of Placebo Controls in the Evaluation of Surgery: Systematic Review." *British Medical Journal* 348: g3253.

Werner, H., and S. Marien. 2022. "Process vs. Outcome? How to Evaluate the Effects of Participatory Processes on Legitimacy Perceptions." *British Journal of Political Science* 51 (1): 429–36.

Wildavsky, A. 1979. *Speaking the Truth to Power*. Boston: Little Brown & Co.

White, H. 2019. "The Twenty-First Century Experimenting Society: The Four Waves of the Evidence Revolution." *Palgrave Communications* 5 (1): 1–7.