

Georgetown University

From the Selected Works of Karl Widerquist

February, 2018

[First Draft] The Devil's in the
Caveats: A Critical Analysis of Basic
Income Experiments for
Researchers, Policymakers, and
Citizens

Karl Widerquist



Available at: <https://works.bepress.com/widerquist/86/>

The Devil's in the Caveats:
A Critical Analysis of Basic Income Experiments for Researchers,
Policymakers, and Citizens

Karl Widerquist, Georgetown University-Qatar

Karl Widerquist

Draft prepared for the Center for International and Regional Studies' (CIRS) book workshop,
Georgetown University-Qatar, March 26, 2018

I would like to give special thanks to Zohaib Tahir, my research assistant, who provided incredibly useful help. I would also like to thank Misba Bhatti, Zahra Barbar, Mehran Kamrava, and everyone at CIRS for funding and organizing this workshop, for funding my research assistant, for showing so much confidence in me, and for giving me a deadline to finish this thing quickly. Thanks to the ten people who will be coming to the workshop. Thanks to Kate McFarland and Karsten Lieberkind for useful comments on an earlier draft. Thanks to everyone who attended my various presentations of these ideas and gave me feedback, especially in the two small-group presentations I made at Georgetown-Qatar.

Contents

Part one: UBI, available tests, testing problems, and past experiments

Chapter 1: Introduction

Chapter 2: Universal Basic Income and its more testable sibling, the Negative Income Tax

Chapter 3: Available testing techniques

Chapter 4: Testing difficulties

Chapter 5: The practical impossibility of testing UBI

Chapter 6: BIG experiments of the 1970s and the public reaction to them

Chapter 7: New experimental findings 2009-2013

Part Two: The place of experiments in the political economy of UBI

Chapter 8: Why UBI experiments cannot resolve much of the public disagreement about UBI

Chapter 9: The political economy of the decision to have a UBI experiment

Chapter 10: The chain of misunderstanding between experimenters and their nonspecialist audience

Chapter 11: Overcoming spin, sensationalism misunderstanding, and the streetlight effect

Part Three: From the debate to the test

Chapter 12: The bottom line

Chapter 13: Identifying important empirical claims in the UBI debate

Chapter 14: Claims that don't need a test

Chapter 15: Claims that can't be tested with available techniques

Chapter 16: Claims that can be tested but only partially, indirectly, or inconclusively

Chapter 17: From the dream test to good tests within feasible budgets

Part one: UBI, available tests, testing problems, and past experiments

Chapter 1: Introduction

The devil's in the details is a common saying about policy *proposals*. Perhaps we need a similar saying for policy *research*, something like *the devil's in the caveats*. I say this both because nonspecialists (citizens and policymakers who are ultimately responsible for evaluating policy in any democracy) have great difficulty understanding what research implies about policy and because specialists often have difficulty understanding what citizens and policymakers most hope to learn from policy research.

This problem creates a great difficulty for Universal Basic Income (UBI) experiments which are now getting underway in several countries. They can provide useful evidence, but the best we can hope is that these experiments will add a small part to the existing body of evidence people need to fully evaluate UBI as a policy proposal. Specialists can provide caveats about the limits of what research implies, but nonspecialists are often unable to translate caveats into a firm grasp of what that research does and does not imply about the policy at issue. Therefore, even the best scientific policy research can leave nonspecialists with an oversimplified, or simply wrong, impression of its implications for policy.

For example, popular media reports about medical research often leave people with the impression that medical consensus about prevention and treatment of diseases swings widely from one recommendation to another, when medical consensus is actually slow to change and even slower to reverse a change once made. This misconception exists in part because most nonspecialists don't have the background to understand how definitive or reliable one piece of research is or to understand the difference between a medical consensus and an oversimplified or sensationalized popular media report about one study.

Whatever problems medical research has with misunderstandings of this type, social science research in general—and UBI experiments in particular—has greater problems. At least some medical questions can be adequately addressed in a simple individualized, controlled experiment. But as this book will show, UBI has many complex economic, political, social, and cultural effects that cannot be observed in any small-scale, controlled experiment. Even the best experiment contributes only a small addition to a large body of knowledge about UBI and leaves many important questions unanswered.

Experimental evidence alone cannot answer the big questions about UBI, such as does it work as intended; should we introduce it on a national level? The difference between what an experiment can show and the answers to big questions like these are enormous. When research is conducted of, by, and for specialists, mutual understanding of the limits of research usually requires no more than a simple list of caveats, many of which can go without mention in a group with a great deal of shared, specialized knowledge.

The same is not true when policymakers and citizens are an important part of the audience—as they are for research on major policy issues such as UBI. Citizens and policymakers understandably try to interpret experimental results in as answers to the big questions.

Unfortunately, the difference between what experiments show and what people most want to know is difficult to understand, and lists of caveats are of it are often tedious to read and difficult to interpret. Therefore, citizens and policymakers have great difficulty understanding what UBI experiments do and do not imply about those big questions.

This book discusses the difficulty of conducting UBI experiments and communicating their results to nonspecialists given both the inherent limits of experimental techniques and the many barriers that make it difficult for researchers, journalists, policymakers, and citizens to understand each other. Policymakers, citizens, and journalists with a better understanding of these barriers will be better able to understand and apply the information they learn from experiments. Specialist with a better understanding of these barriers will not only do a better job reporting their research; they will also be able to design and conduct experiments that are more useful to the policymaking process.

The book assumes that one important goal of UBI experiments is or ought to be to raise the level of debate among citizens and policy makers by expanding their understanding of evidence. The vast majority of research specialists who conduct experiments are not fools or fakers; they will find evidence that makes a positive contribution to the body of knowledge about UBI. But that contribution will be a small part of what we want to know, and the effort to translate that contribution into a better public understanding of the body of evidence about UBI is far more difficult than the effort than usually recognized.

Unfortunately, although research specialists are professionals at communicating with other specialists, the vast majority of them are amateurs at communicating with nonspecialists—and I am no exception. Very often specialists don't know what evidence would be most valuable to citizens or policymakers, while citizens and policymakers don't understand the value of the evidence that researchers are able to find. The effort to raise the level of political debate requires knowledge and skills that researchers have no special training to do and creates risks that less politically tied research does not have, including the vulnerability to spin, misuse, sensationalism, or oversimplification. The risks are so great that past UBI-related experiments—despite almost always being good science—have a mixed record at raising the level of debate among nonspecialists. Some have. Some have not.

The devil *is* in the caveats. As we begin a new round of experiments, it's important to consider lessons in how to improve the chances that experiments will successfully enlighten the public discussion of UBI.

Researchers might be tempted to ignore this issue, to produce research for researchers or for whoever commissioned the study, while considering the relationship between the experiments and the political process to be out of their purview. Unfortunately, UBI experiments are too closely tied to the political process and their results are too easily misunderstood for researchers to ignore the experiments role in the political economy of the UBI debate.

Although UBI experiments are scientific endeavors, they are both an outcome of and an input into the political process. The current experiments are—directly or indirectly—a response to the growth of the UBI movement in the years leading up to the sudden spread of interest in experiments. UBI or similar experiments are enormous undertakings—generally too big to be funded by a simple grant from a science foundation. The 1970s experiments were commissioned, not by science foundations, but by acts of national legislatures that were seriously considering the policy. The same is true for the new government-funded experiments, such as those in Finland and Canada. The experiments in Namibia, India, Kenya, and California are all funded by private organizations with a strong interest in the UBI debate.

Even if policymakers who commission UBI experiments and researchers who conduct them are most interested in a few narrowly focused technical questions, they ignore the place of those experiments in the UBI debate at their peril. Whether researchers like it or not, people on all sides of the UBI debate all over the world will look to UBI experiments for information about UBI and sometimes for ammunition to use in that debate. The experiments will affect the public discussion UBI. People will seize on finding and say it implies X about whether UBI is cost-effective or whether we should introduce it. The data will be used this way. The question is whether it will be understood and used well or misunderstood and abused.

The goal of this book is to help researchers, policymakers, citizens, journalists, and anyone else interested in UBI experiments understand the limits of UBI experiments and bridge the gaps in understanding between them to help the experiments succeed in the goal of raising the level of debate. Policymakers, journalists, and citizens who understand the place of experiments in the political economy of the UBI debate will get more out of the experiments and be better equipped to counter spin, misuse, sensationalism, or oversimplification of experimental findings.

Researchers who understand the place of experiments in the political economy of the UBI discussion can obviously communicate their results more effectively. But equally or perhaps even more important, research who understand and respect the public discussion can design better experiments. Although the belief that experiments can answer the biggest questions about UBI is naïve, the desire for an answer is not. Therefore, I suggest that researchers keep in mind what I call the bottom line: an overall evaluation of UBI as a long-term, national policy.¹ Experiments alone cannot provide enough evidence to answer a bottom-line question, but they can research claims that people care about and they can relate all of their findings to the bottom line. Virtually all research into UBI is important to the extent that it helps people make that bottom-line evaluation, but it is extremely difficult to translate experimental findings into clear implications for the bottom line. Why the bottom line is so important and how to focus on it are major subjects of this book.

To achieve the goal of bridging communication gaps, this book has to focus extensively on how limited UBI experiments are in answering the big questions about UBI. And it has to explain the many significant barriers there are to communicating their results in a way that successfully raises the level of debate. Therefore, this book has a lot of negative things to say, but that should not distract readers from my overall enthusiasm for UBI experiments. They are worth doing. They are worth doing well. And they are worth the effort it takes to understand what they are and are not capable of contributing to the public decision of whether to implement UBI.

Despite my enthusiasm, some readers might react to all of the difficulties this book discusses by giving up on experiments. Some readers might be unenthusiastic for other reasons. To these readers I say, experiments are coming; it's important to make the best of them. We make the best of them when specialists and nonspecialists understand each other better.

The book's goal is not a negative one. It's not simply about how an experiment can fail to raise the level of public discussion. It's about what everyone involved can do to help it succeed.

With these goals in mind, the book is divided into three parts. Part one includes seven chapters providing necessary background and discussing the history of UBI and related experiments. After the current introductory chapter, Chapter 2 explains what UBI is and how it works. It also explains the workings of UBI's more testable cousin, the Negative Income Tax (NIT), leaving the explanation of why NIT is more testable for later.

Chapter 3 discusses some necessary definitions and the pros and cons of the available testing techniques: randomized controlled trials (RCTs), saturation studies, and combinations of the two.

Chapter 4 discusses several general problems that virtually any UBI experiment will have to deal with: the virtually impossibility of observing community effects, the difficulty of predicting long-term effects, the Hawthorne effect, the difficulty in separating the effects of the size and type of program being studied, and the streetlight effect.

Chapter 5 discusses one big difficulty: the impossibility of testing UBI under most practical circumstances and the problems creating by using NIT as an approximation of UBI.

Chapter 6 discusses the five NIT experiments conducted in the 1970s in the United States in Canada, summarizes their findings, and shows how badly those findings were misunderstood in the public debate at the time. It argues that although the experiments succeeded in the goal of obtaining useful data, they badly failed in the goal of raising the level of debate at the time.

Chapter 7 discusses more recent experimental findings from two experiments conducted in the late 2000s and early 2010s and from newly released data from one of the 1970s experiments. It argues that these findings had a much more positive impact on the level of debate.

Part Two focuses on the role of experiments in the political economy of the UBI debate. It discusses why specialists and nonspecialists have such a hard time understanding each other's perspectives on UBI and UBI experiments, what kind of problems this misunderstanding can cause, and how to reduce it.

Chapter 8 explains why UBI experiments cannot resolve much of the public disagreement about UBI. It argues a great deal of evidence is already available. Experiments can only make a small contribution to that body of evidence, and the debate turns crucial not on remaining unknowns about UBI's effects but on the controversy over the ethical desirability of UBI and its known effects.

Chapter 9 discusses the political economy of the strategic decision to have a UBI experiment in light of its limits. It argues that experiments are a risky strategy for the UBI movement because of their vulnerability to misunderstanding and misuse.

Chapter 10 examines why the results of experiments are so easily misunderstood, and therefore, vulnerable to misuse. These problems happen because of the inherent complexity of the material and the differences in background knowledge of the people involved. The nonspecialist who create the demand for experiments and who need to understand the results to make public policy decisions are separated from the specialists who conduct the experiments by a long and difficult communication chain.

Chapter 11 discusses how to overcome these communication barriers and reduce the problems associated with them. It argues that it is not enough to communicate the findings of experiments on their own terms, but people also need to be made aware of the role experiments play in the political economy of the UBI debate.

Part Three discusses conducting a UBI experiment in a way capable of raising the level of discussion in light both of the difficult political economy outlined in Part Two and of the inherent difficulty of reproducing the relevant aspects of UBI on an experimental level.

Chapter 12 begins this effort by suggesting that UBI experiments should relate all findings to what I call the "bottom-line question," the overall cost-effectiveness of a fully implemented national UBI. It also suggests possible wording of a cost-effectiveness question.

Chapter 13 proposes a list of important empirical claims made by supporters and opponents of UBI in an effort to identify what empirical questions are important to the public discussion of UBI.

The next three chapters consider how well experiments can address those claims. Chapter 14 considers several claims that can be analytically shown to be true without the need for a test.

Chapter 15 identifies several empirical claims that should not be ignored but that cannot be tested on an experimental scale. Evidence about these claims will have to come from other sources, which will have to be combined with experimental evidence to connect it to the bottom line.

Chapter 16 identifies several claims that can be tested on an experimental scale, but shows how each one can only be tested partially, indirectly, and/or inconclusively. It suggests research questions to address the issues and how to relate experimental findings to the questions people most want answered.

Chapter 17 concludes with a discussion of possible ways to test UBI in light of these issues, working down from the dream test that solves all testing problems to tests that might be possible within the experiment's budget.

Chapter 2: Universal Basic Income and its more testable sibling, the Negative Income Tax

UBI is commonly defined as a periodic, cash income paid individually to all members of a political community without means test or work requirement.² It is also commonly understood to be regular, stable in size, and lifelong, although it might be lower for children or higher for people of retirement age. Although this definition probably reflects the most common usage of the term, UBI is a contested concept that is used differently in different political contexts and by different people in the same context.

Some people add the criteria that the UBI must be high enough to achieve some criteria of sufficiency. UBI experiments have tended to focus on versions of UBI meeting *some* criteria of sufficiency, not necessarily one that all UBI supporters would view as sufficient.

Some people subtract the criteria that the UBI is paid individually and without a means test. That is, it could be paid to a household and phased out as (household) income rises. Negative Income Tax (NIT) is the more common name for a program that lacks those two criteria but otherwise qualifies as a UBI. The defining difference between the two is that NIT's the grant is reduced as income rises. The second characteristic (that it is given on a household rather than an individual basis) follows from the first because most households pool their income and pay taxes as a unit.

Not everyone recognizes this difference between NIT and UBI. For example, in Canadian, where one of the current UBI experiments is happening, the terms UBI and NIT are often used equivalently and the NIT version, under the name "basic income," currently dominates the discussion among policymakers.

The NIT is important to any discussion of UBI experiments because the differences between NIT and UBI make NIT more easily testable in an experiment. Therefore, this section explains both of them.

Under the common definition of UBI, every citizen of a nation (or every legal resident of a region) receives a regular income from the government (or some other authority) regardless of whether they have any other income, wealth, potential for employment, and so on. That's all there is to UBI in the definitional sense, but it has an additional inherent feature not reflected in its definition but necessary for its operation: UBI has to be financed with taxes or it will cause rampant inflation. It is conceivable that UBI could be financed by some enormous jointly-owned asset but the revenues of that asset could be considered to be equivalent to taxes, and in most political contexts such an asset could not even be created without introducing large new taxes, and so this book focuses on the tax-financed model.

Any UBI is defined by two essential parameters: the “grant” level (G), which is simply the size of the UBI, and the “marginal tax rate” (t), which is the rate at which taxes gradually become larger than the UBI. Any tax could be used to support UBI. Popular options include income tax, wealth tax, sales tax, and resource tax (i.e. taxes on the rental value of privately owned natural or socially created resources such as land, the broadcast spectrum, and the banking system). Given the need to finance UBI, the actual financial benefit any individual gets from the UBI system is its *net* benefit, given by the following equation:

$$\text{Net Benefit} = \text{UBI} - \text{Taxes}$$

Or:

$$\text{Net Benefit} = G - t \times (\text{the tax base})$$

For an income-tax-financed UBI, this equation becomes:

$$\text{Net Benefit} = G - t \times (\text{income})$$

The income-tax-financed UBI is not necessarily the most popular version of the program, but it simplifies the mathematics and is, therefore, popular with researchers.

The NIT pays a regular, cash income to all members of a political community without a work requirement, but it has a means test; it is gradually phased out as income rises. NIT is so similar to an income-tax-financed UBI that the same formula can be used to show how the net benefit of both:

$$\text{Net Benefit} = G - t \times (\text{income})$$

The difference is that under UBI, G stays the same as taxes increase while under NIT, taxes remain zero as the grant (i.e. the “negative tax”) is gradually reduced to zero—at the breakeven point—and only then are taxes (i.e. positive taxes) introduced.

Some people argue that NIT and UBI are effectively the same policy with insignificant administrative differences. But others argue that the differences are important. Some differences are purely administrative: the NIT saves the trouble of paying a UBI to net contributors and taking it back from the same people in taxes, but the UBI saves the trouble of determining who is eligible at a moment's notice when someone suddenly loses their income. Presumably, people will have to apply for an NIT and prove that their income has gone down before they receive it. This process could be difficult for people in a sudden economic crisis, such as a divorce, the loss of job, or the

failure of a business. No such issue exists with UBI. It would be directly deposited into one's account regardless of whether taxes were also coming out of one's paycheck. As an individual grant UBI might make it harder for one spouse to dominate the family's income.

Terms such as "Basic Income Guarantee" (BIG) and "Guaranteed Income" are sometimes used generically as terms for things that fit the definition of either UBI or NIT. BIG ensures that everyone has a nonzero income whether or not they have any private income. Either form of BIG can be used to maintain the same minimum guarantee level for people who have no other income. This assurance of an unconditional non-zero income is the essential characteristic of both policies and the heart of what separates UBI from conventional policies.

The controversial question among supporters is whether the seemingly small administrative differences between the two policies are significant enough that one model should be preferred over another. This is a question that one would ideally want to address in a test, but later chapters will show, tests usually have to focus on NIT.

Chapter 3: Available testing techniques

After this chapter defines some relevant terms, it discusses the pros and cons of the techniques available for testing UBI.

All empirical research (whether experimental or not) attempts to answer question called appropriately the research question. Often a large study, like a UBI experiment, will ask a serious of research questions. A question like, "what are UBI's effects," is too vague to be a useful research question. A UBI could have an infinite number of effects, some important some trivial. Although researchers would happy to discover effects they were not looking for, you can find an effect that you make no effort to measure.

Therefore, most research questions are formulated around hypothesis testing. That is, they test a claim about a supposed relationship. For example, a lot of medical research tests the hypothesis that a medical treatment is safe. The research question, "is it safe," might still be too vague, and might need to be replaced by a more testable research question like, "is it correlated with increased diagnosable health problems."

Empirical studies cannot always conclusively verify or falsify any claim, they can only state whether the evidence is consistent with or contradictory toward the claim, but this is often enough to extremely useful.

Sometimes the hypothesis that a treatment will have some effect will not be in serious doubt, and the research question will become, "How large will this effect be?" That sort of a research question is useful to examine, but it is not a hypothesis test unless it is paired with a hypothesis that the effect will be larger or smaller than X amount. For example, in wealthy countries, past evidence indicates UBI will correspond with a decline in the average time recipients spend in employment. The question is how much will it decline. What size of a finding would be significant? Is it that the response is greater than zero? If so, we don't need a test. Is it that the response is greater than X%? If so, among which group? Is that it is large enough to make the program unsustainable? Or is it something else entirely: perhaps the significance of this response is no in how large it is, but in some qualitative measure of what people do with the reduced time they spending working? These different hypotheses that research findings can be used to test create great problems later in the book.

Two desirable attributes for estimates are that they are "accurate" and "unbiased." An accurate estimate is one that is likely to be close to actual value. An "unbiased" estimate is one

that is just as likely to overestimate the actual value as it is to underestimate the actual value. That is, it lacks “statistical bias.” Statistical bias is very different from the bias in the sense of being favorability to one group or another.

Statistical bias cannot always be eliminated, and sometimes it has to be traded off against accuracy. A slightly biased estimation technique could be preferable to an unbiased but less accurate measure. For example, suppose you were estimating a person’s age. The first technique is likely to produce results within the range of 0 to 2 years over the person’s actual age. The second technique is likely to produce an estimate within the range of 20 years below to 20 years over the person’s actual age. The first technique is biased, but it is so much more accurate than the second, unbiased, technique, it’s almost certainly more useful.

Bias causes great difficulty for empirical studies. Sometimes you don’t know whether a technique is biased or not. Sometimes you know that it is likely to be biased, but you don’t know the direction of the bias. Sometimes you know that it is likely to be biased in a particular direction, but can’t simply counteract the bias because you don’t know how biased it will be. Later chapters show that all of these biases affect the techniques available for testing UBI.

Statistical bias is not the only sense of “bias” used in this book, but it is the primary one, and I think it is clear from context which sense the book uses at any time.

One surprisingly controversial definition issue is what to call the effort to try out UBI on a small scale to learn something about it in advance of full implementation. In common English, the words “test,” “trial,” “pilot,” and “experiment” all fit that broad definition, but some of them are also used in more specific senses in some circumstances—often in technical settings.

“Experiment” is sometimes used to refer only to a “randomized controlled trial” (RCT): a test designed to isolate the effects of the factors being studied by using randomization as a method to control as much as possible for all other factors that might influence the relevant outcomes. Researchers do so by randomly selecting two sufficiently large groups that differ as little as possible from each other and from the wider population. They give the treatment to one group only (the experimental group), and observe whether that group differs in relevant ways from the other group (the control group). If the groups are sufficiently large and properly selected, the differences between them—other than those caused by the treatment—will tend to cancel each other out. This method is indispensable in many forms of medical research, and it can be useful in social science as well. But as argued below, it is not always the best way to address questions at issue in the UBI debate.

“Pilot” or “pilot project,” can be used as a broader alternative to “experiment,” but it carries baggage as well. “Pilot project” sometimes implies that the test is conducted by an authority with the power to fully implement the policy—at least if the pilot meets some criteria of success. Sometimes it implies that a firm decision in favor of full implementation has already been made, and the test is being used to determine *how* rather than whether to implement it.

Even the simple word “test” sometimes implies that the study involves some firm criteria by which the policy will be judged to have passed or failed. Nonspecialists often expect such criteria from experiments of any kind. Social science experiments are usually conducted without any criteria of success in mind and often in a context in which the researchers conducting the experiment cannot impose one without imposing their own values on difficult and controversial political debates. Therefore, it’s best to fight this impression as much as possible.

The term, “trial” or “implementation trial,” has the fewest other connotations, and so I occasionally use it for clarity, but it is also the least familiar of all of these terms.

I mostly use the term “experiment” in that broader sense defined in the first paragraph of this chapter, despite how, as explained below, at least some specialists assert the common usage is wrong.

What distinguishes an experiment, test, trial, or pilot in this broad sense from a non-experiment is that an experiment is in place solely (or at least primarily) to learn something about the policy. It is not (primarily) an attempt to *implement* the policy. In this sense, the NIT experiments of the 1970s were trials, but the Alaska Dividend and Cherokee per capita payments (for example) are not. Although these policies might provide useful opportunities to learn something about UBI, they are not put in place for that opportunity. They are full-fledged policies that were in place in hopes of achieving goals.

Once the decision is made to conduct an experiment, researchers have a choice of two broad types of techniques or a combination of the two. The first, an RCT, is defined above. The second, a “saturation study,” involves identifying two relevant communities, such as two small towns, giving the treatment to everyone in one community and not to people in the other. Although researchers might randomly choose which of the two sites will be the control and which the experimental site, that level of randomness is not enough to control for other factors that might make one site different from another. Although the communities could be selected to be as similar to each other and to the wider population in as many observed ways as possible, they might differ in important but unobserved ways.

Both RCTs and saturation studies are useful. RCTs are better at examining issues in which most of the effects occur at the individual level. Saturation studies are better at examining issues in which many important effects occur at the community level. These “community effects” are extremely important for UBI because its effects depend on the interactions of people in markets and cultural settings.

Whether the trial is an RCT or a saturation study, the experimental and control groups each need to be at least a few hundred (and preferably a few thousand people) to produce statistically useful results. Once the experimental group begins receiving “the treatment,” researchers observe how they behave in comparison to the control group. The central goal of any experiment is to find a way to ensure that differences between the two groups will be attributable as much as possible to the treatment and to random fluctuations, which will mostly cancel out in a large enough sample. Hence the *control* in the experiment. Unfortunately, in social science, creating a trial that is both controlled and representative of how the policy under investigation will work under full implementation is extremely difficult.

Some researchers—labeled “Randomistas” by their critics—insist that only an RCT is truly scientific or truly deserving of the term “experiment.”³ One reason to resist the Randomista use of “experiment” is to avoid confusion caused by the belief that more technical definitions are the “right” definitions. That is not how language works. Specialists do not own the language. The most commonly used definition is the most acceptable definition. Specialists who insist that technical definitions are the only right definitions risk confusing nonspecialists who will be most familiar with the common understanding of “experiment” and who are important consumers of the findings of UBI experiments—or any policy-related experiment.

Another reason to resist the Randomista use of the word is that RCTs are not accurately described as the only scientific form of experiment. RCTs make some valuable statistical techniques available that aren’t available with saturation studies, and they make it possible to control for unobserved factors that saturation studies cannot conclusively control for. But they do so by entirely ignoring certain kinds of effects (discussed below). In other words, RCTs control

for more things but test fewer things. Therefore, researchers should be open to using both RCTs and saturation studies as appropriate. Both techniques should be considered part of the social scientists' toolkit as long as researchers are careful to note the extent to which their results should be seen as tentative or conclusive and the ways in which those results are likely to be biased.

Both techniques have some advantages over the other in each of these respects, important effects of UBI occur at both the individual and the community level. Individuals immediately react to UBI in many important ways that are worth estimating, but they interact with other individuals in markets, society, culture, and politics. All of these interactions generate important feedback effects throughout the community. Existing theory and empirical evidence indicates that some community effects might be as important or more important than the initial individual effects of UBI. If researchers opt only for an RCT (perhaps so that they will be taken seriously by Randomistas), they must choose between ignoring feedback effects entirely or supplementing their experimental data with information from other sources to simulate feedback. Guy Standing argues that the Randomista attitude often leads to ignoring community effects even on issues—such as UBI—where such effects are likely to be extremely important.⁴

Because both types of experiments have advantages and disadvantages, an ideal test would fully combine saturation and RCT techniques by randomly select dozens of saturation sites for both the control and experimental groups. For example, consider a test of whether a vaccine creates “herd immunity,” which refers to the way a large number of individuals with immunity in a group helps protect individuals without immunity—perhaps because they are not individually responsive to vaccines. The individual immunity question can be answered by a simple RCT with a few hundred or a few thousand subjects. The questions of how large or isolated a herd needs to be to establish herd immunity might require dozens or even hundreds of herds of varying sizes and levels of isolation to get statistically significant results. For herds of livestock, such a test might be affordable. For herds of humans, it is very likely to be unaffordable.

Researchers have conducted experiments with multiple saturation sites in India and Kenya where poverty is extremely high and a UBI of a dollar a day or less is extremely significant to study participants. Unfortunately, the cost of a saturation site large enough to capture all of the relevant community effects or a statistically significant number of saturation sites—enough to satisfy the Randomistas—is usually prohibitively expensive even in less wealthy countries, and it will almost certainly be prohibitively expensive in higher-income countries (more on these issues in the next chapter).

Most likely, in wealthier countries, the techniques available will be limited to one RCT or one saturation site, or perhaps a combination of the two, but not much more than that.

Chapter 4: Testing difficulties

This chapter discusses several general difficulties that are likely to affect any UBI experiment. For each one, it considers general ways of dealing with these problems. Chapter 5 continues with a more thorough discussion of one big general problem for UBI testing. Part Three of this book discusses the problems these issues present for testing the specific claims of interest in the UBI discussion.

Community effects

Community effects, defined in the last chapter, provide an enormous difficulty for UBI experiments even if the experiment combines RCT and saturation techniques, because UBI is likely to have a lot of them; we cannot know in advance what all of them might be; we don't know the size and level of isolation of a community necessary to bring out various different community effects UBI might, and most importantly some of UBI's community effects probably happen at the national level.

The community effects that exaggerate individual effects are the easiest to understand, for example, evidence indicates that children who are less poor are healthier and children who live in a less poor neighborhood and less poor countries are also healthier.⁵ Confirmation from an RCT that UBI has this effect on childhood health at the individual level can be extrapolated to indicate the possibility that the effect will be greater at the national level—at least if the theory used to make that extrapolation is well-explained.

Community effects that play out very differently than individual effects are much harder to extrapolate from individual data. For example, some obvious and important community effects of UBI have to do with the feedback effects between workers and employers. Workers (at least in wealthier nations) are likely to respond to UBI by working less. Employers are likely to respond to that action by workers by offering better wages and working conditions. Workers are likely to respond better wages and working conditions by working more, partially counteracting their initial drop in hours worked. Call that a feedback loop. Culture, education, and other factors are likely to respond to the changed labor market, and these factors could feedback to other labor market changes. That is a feedback loop with five potential steps. An RCT can measure only the first step in the six steps in that predicted loop. A saturation study might capture some of the second and third steps, but only to the extent that these effects occur at the local level.

Ideally, the existence of feedback loops like this is something we would like an experiment to investigate. To do so, we would need something like the herd immunity test described in Chapter 3, but any such test will almost certainly be prohibitively expensive. The affordable options in wealthy countries are likely to be no more than one small, saturation since and/or on RCT. Clearly, effects that occur only at the national level cannot be observed before introducing UBI at the national level. But some important community effects might occur at levels as small as the neighborhoods. The ghettoization of poverty is said to affect health, crime, education, and many other important issues.

Researchers with limited budgets have at least four options for dealing with community effects. Each of them has a serious downside. First, conduct an RCT only and ignore community effects entirely. This option would bias the results, sometimes in unpredictable directions, and even if the direction of bias is predictable, the size of the bias is usually not.

Second, conduct an RCT only, leaving all those biases in place, but include caveats explaining them. This option is likely to be the most popular among researchers, but it has a large potential downside. Specialists often don't always explain caveats well. Nonspecialists find them tedious; they often ignore or don't understand them, and caveats often get lost in the chain of communication connecting specialists to citizens and policymakers. In practice, this second option might not be that different from the first option. The 1970s U.S. experiments might have attempted this option, but as Chapter 6 shows, the public discussion proceeded with little or no recognition that community effects existed or were being left out of the study.

Third, conduct an RCT and use theory and data from other sources in computer simulations to estimate community effects. This option means the report on the experimental findings will be driven less by those experimental findings and more by the assumptions of that simulation model. Hopefully, many of the assumptions of those simulation models will be drawn from very good evidence, but evidence of the quality we would want might not always be available. This option does not eliminate the need for caveats, it merely changes the nature of the caveats involved. Instead of explaining the why experimental findings don't answer the questions people most want answered, the caveats would have to explain three other things (1) the limited role the experimental findings played in the answers given to those questions; (2) the sources, quality, and role of other the evidence used; and (3) the techniques employed to combine the evidence. These caveats risk being misunderstanding just like the original caveats.

Fourth, conduct a saturation study on at least one site (more if budget allows), combined (if budget allows) with an individualized RCT at another site or across a wide geographical area. Small, isolated communities are more likely to show the kind of community effects we can expect at the national level. For example, local businesses have to draw labor from potential employees who are all eligible for UBI. Labor markets are in many ways national and so even a saturation study is likely to be biased toward underestimating employer response, but they are an improvement on RCTs, which are biased toward estimating no response whatsoever.

Unfortunately, a saturation study won't provide evidence about how similar the community effects at the saturation site will be to the community effects of a national program. And individuals in smaller, more isolated communities might not be representative of the people in larger, less isolated communities where the majority of the world's population lives. This problem will bias the study in unknown ways. Any community participating in such a study is likely to be the subject of a great deal of media and social-media attention, which might affect participants' behavior (see discussion of the Hawthorne effect below). If researchers are able to combine an individualized RCT with a saturation study on at least one site, they can compare results from both studies and from simulations they might run to try to get a sense of which differences come from the biases of each technique and which come from community effects.

Manitoba's "Mincome" experiment in the 1970s combined a saturation study with an RCT, and it made a unique contribution to the experiments conducted at that time (see Chapter 6 below). But so far, it remains the only UBI saturation study yet conducted in a higher-income nation. With all the UBI experiments currently getting underway around the world, researchers have the possibility of experimenting with very different techniques. It would be a shame if not one of the experiments in higher-income nations included a saturation site.

Each method is imperfect, but as this book stresses throughout, any social science experiment has a great deal of imperfections. That fact is no reason to give up learning all we can from whatever type of test we are able to conduct. But it is a reason for everyone interested in its results to understand and be content with the experiment's small contribution to the great body of existing-but-still-incomplete knowledge about UBI. The task of making policymakers and citizens understand the relevant implications of those imperfections is extremely difficult. Once again, the devil's in the caveats.

The Hawthorne effect

The "Hawthorne effect" refers to the way observation affects behavior. People in an experiment know they're being observed, and this knowledge might affect their behavior in

unpredictable ways, causing many different forms of bias. Perhaps seeking approval of the observers, participants would behave in ways they think will make them look good or smart or successful to the observers. Perhaps instead they would show off, trying to be funny or interesting or trying to cultivate some kind of image. Perhaps they would try to “help” the observer by displaying what they think the observer wants to see. Perhaps they would try to “harm” the observer by displaying the opposite of what they think the observer wants to see, possibly because of some antagonistic feelings toward either the researcher or the reach objective. Perhaps they would be affected by the power of suggestion: knowing that the observer wants to know whether they do X might unconsciously make them do X or make them avoid doing X more than they normally would. These reactions might sound silly, but no one can completely free themselves from them. Hawthorne effects have been recognized for decades, but exactly how they are likely to affect research remains a mystery,⁶ making it very difficult to compensate for them.

Hawthorne effects are likely to be a bigger problem for the new round of UBI experiments than they were for the 1970s experiments. People are far more subject to observation than they were in the 1970s. Most people post about themselves on social media, and it will be difficult to get them avoid posting about a trial they are participating in. This visibility will make it easier for the media to find them.

Saturation studies in the internet-connected world will be the biggest problem with the Hawthorne effect. A saturation site cannot be kept secret. Participants might find journalists, bloggers, activists, and long-lost friends contacting them on social media to ask what it’s like to be in the UBI saturation study.⁷ How this increased attention will affect their behavior is unknown. I hope it does not make it impossible to do saturation studies in well-wired countries, but it might.

Long-term effects

Any experiment is going to be short-term compared to how long the actual policy is likely to stay in place, and short-term effects might differ significantly from long-term effects. This problem is one that is intuitively easy to grasp for people with no special training, but still, it creates potential problems for understanding research. If some medicine was invented a year ago, no matter how extensively it’s been tested, we can’t rule out the possibility it has some side effect that only shows up ten years later. How much we should worry about that problem depends whether the medicine is a pill a patient would take once or one patient will take every day for the rest of their lives. Extrapolating from past knowledge, medical researchers can do something to get an idea of long-term effects. For example, they can look to see whether it causes known toxins to build in the body, but without longer-term observation, some long-term effects will remain uncertain.

UBI (or any other big public policy program) is more like the pill people take every day for the rest of their lives than the one-time treatment. It’s a long-term policy, and we most want to understand its final, overall, long-term effects. These effects are likely to be community effects that develop out of economic and cultural interactions between people over a very long period. But experiments directly observe only the initial steps in that long, complex chain of reactions.

Researchers can try running a longer-term experiment, but doing so increases the expense and the time it takes to get results, and so most studies are very short-term. The SIME/DIME study contained the longest-run observations so far. It was originally planned for six years. After about three years, researchers obtained permission to extend the experiment to 20 years for a small sub-sample, but that effort was cancelled after nine years.⁸ That is, a small group was eligible for an

NIT for nine years, and less than six years of that time, participants were lead to believe they would receive the NIT for 20 years. Researchers did not find major differences between this group and the shorter-term sample, but this RCT had no way to measure community effects, which are likely to be larger in the long run. Whether a national UBI would work differently in the long term still remains questionable.

The effects of UBI on health, education, labor time, wages, working conditions, and so on might play out over a very long time. Many long-term effects are likely to be in the same direction (at least) as short-term effects. Other long-term effects might partially or fully reverse in the long term. And the best we can do is extrapolate based on theory and data from other sources, imposing yet more assumptions about things we would rather like to learn from an experiment.

The difficulty of separating the effects of the size from the effects of the type of policy being studied

UBI would be an enormous change in social welfare strategy, and if introduced as envisioned by most of its supporters, it would also involve a large increase in social welfare spending. If we want a study to help us to understand how UBI works, how its effects differ from other policies, and whether money is better spent on UBI or other policies, we need to separate the effects of the size of the program being studied from the effects of type of program being studied.

Separating the effects of size and type is extremely difficult in a UBI experiment. The experiments in the United States in the 1970s tested various sizes of NIT, but they only had one control group, all the members of which were eligible for the welfare system existing at the time (see Chapter 6). Thus, the effects of the larger NITs were compared to the effects of the existing system and to smaller NITs, but not to equally generous versions of the existing system. This method gave some information about the how the effects of NIT differ by size and some idea about how the effects of NIT differed from the effects of the existing system, but it could not determine the extent to which the effects of the larger NITs had more to do with their being larger or more to do with their being a different type of system.

Furthermore, most of the reporting of results (including those summarized in Chapter 4) lumped together the findings from the various experimental groups with various grant levels and marginal tax rates. This amalgamation not only makes it difficult to separate the effects of size and type, but also it makes it difficult to interpret just what size of UBI is being tested on average. What then do the numbers say about the choice between introducing a generous UBI or using the same amount of money to make the existing system more generous or to introduce some other strategy? Not a lot, unfortunately. And that question is far closer to what people most want to know than whether the control group behaved differently from the experimental group. There are two ways to get the estimates closer to what we really want to know.

The first option is to include several different control groups facing differently generous versions of the existing system or whatever system UBI was being tested against. This might seem easy, but to get a really good estimate of the different effects of size and type of spending, each version of UBI would have to be paired with a different strategy of exactly the same size. But unfortunately, for two so different strategies, it's difficult to determine in advance what size is the same. The cost of a public policy depend on overhead costs, take-up rate, and other factors, most of which can't be estimated in an experiment. Researchers can use data from other sources to estimate what an equally-sized version of the existing system might be. Although any estimate will

be highly approximate, just having various sizes for the control groups will help tease out the difference between size and type.

However, none of the NIT or UBI experiments conducted so far have used this technique, and I don't expect any of the currently-under-discussion experiments will either, for one simple reason. It's very expensive. It roughly doubles the cost of the experiment. Researchers will have to give out twice as many checks each week, and they will have to deal with the difficult administrative challenge of determining how much each individual in the control group would be eligible for this week if programs A, B, C, and D were X% more generous. They will have to somehow make up the difference, which is probably difficult enough for cash benefits, and extremely difficult for in-kind benefits such as public housing or Food Stamps.

The second option for examining the difference between size and type is to use theory and data from elsewhere and computer simulations to estimate how the control group would have responded to a more generous version of the existing system and use that as the baseline for comparison or at least as a way to estimate what portions of each observed difference between the control and experimental group are attributable size or type. This method would also be highly approximate, but nevertheless it is a potentially useful check on the simple comparison.

I don't know of any literature on past experiments that attempted to use this method. It certainly was not emphasized in the discussion of any NIT or UBI experiment completed so far. Instead most of the literature reported the observed differences between the control group and the experimental group, mentioning what the two groups were eligible for, and sometimes with no further explanation at all, leaving it up to readers to understand that the results, therefore, involve some amalgamation of the effects of size and type of plan being studied. The popular literature at the time shows little or no awareness of this issue.

The two methods of accounting for it are expensive or difficult or not necessarily very accurate or a mix of all three. Simply explaining the issue takes some effort and all it does is leave readers with the possibly disappointing realization that the numbers are less meaningful than the might initially have appeared.

The streetlight effect

The "streetlight effect" might well be the most important problem for UBI experiments. Don't let the ease with which it can be explained deceive you.

The streetlight effect gets its name from a joke in which a man loses his keys in a dark alley but looks for them under a streetlight because, he explains, "the light is so much better here." In social science, the "streetlight effect" is research focused questions that are easier to answer but less important rather than on questions that are more important but more difficult to answer.

A social science experiment is a tool to help evaluate a potential policy. What's ultimately important about the experiment is its ability to do that. But an experiment is also a very specific tool that is much better at addressing some questions about that policy than others. Even the questions it can address, it can address only in a partial and/or indirect manner—thereby producing information that is importantly different from the most important information for the evaluation of the policy in question.

Experiments will find useful evidence, but to understand its value, requires remaining focused on the big, evaluative questions and making the difficult, sometimes tenuous connection between that evidence and the important questions.

But research reports, academic literature, and popular literature on past experiments have overwhelmingly focused on the differences that experiments were best able to observe between the control and experimental groups as if those observable differences were the most important issues in evaluating UBI and as if understanding the differences between the control and experimental groups could be straightforwardly extrapolated into an understanding of the probable effects of policy once introduced on a national scale. That is, as if there were no community effects, no long-term effects, and so on.

Researchers usually include caveats about those limitations, but the Devil's in those caveats. A simple list of caveats falls far short of a discussion of how the information found relates to the most important questions to ask in evaluating the potential for national adoption of a UBI program.

The potential for the streetlight effect will play a large role in the discussion in Parts Two and Three of this book.

Chapter 5: The practical impossibility of testing UBI

This chapter argues that, in wealthy countries, it is impossible to test UBI in practice: an experiment will either test NIT instead of UBI or it will test a UBI plus an influx of money that would not normally accompany UBI, and the test will be unrepresentative in other important ways. This problem does not mean researchers should give up; experiments can test an approximation of UBI, but understanding and accounting for the biases created by the switch is not easy.

Simulating UBI in a trial might deceptively seem simple: randomly select a group of individuals and give them each a UBI. But as chapter 2 explained, the UBI grant is not all there is to a UBI program. UBI requires taxes, or it will cause rampant inflation. Although everyone gets the UBI, the vast majority of people in wealthy countries also pay at least some taxes. Although the size of UBI is the same for everyone, the net benefit individuals receive varies with the amount of taxes they pay, and the net benefit is what affects their available choices, not the nominal amount of the grant.

While giving out UBI grants is simple, getting the right net benefit for each net recipient in the way UBI gets people their net benefit is not, because researchers can't levy special taxes on participants in an experiment. Researchers have at least three options for dealing with this problem.

The first option is to include in the study only people who would pay little or no taxes under the UBI program being examined. The difference between this group's gross and net benefit from the UBI will be zero or negligible. This solution can work in less wealthy, more unequal countries that have extreme inequality and a large number of very poor people who pay no taxes now and would not need to start paying taxes to finance a significant UBI. The Namibian and Indian experiments studied very impoverished villages where few if any of the residents would pay any taxes at all.

However, in wealthy nations, very few people would pay zero taxes under most UBI schemes, and there would be no stable population of zero-taxpaying people to focus on. Under a reasonably affordable version of UBI, people would probably have to start paying taxes from a very low income, so that their net benefit gradually declines as income rises at a rate that will reach a reasonably affordable break-even point. The taxes don't have to be income taxes, but the tax has to fall partly on net recipients because a UBI scheme that gives a net benefit equal to the full value of a poverty-level-or-higher UBI grant to people with higher-than-poverty private

incomes is not likely to be affordable. For this reason, many UBI schemes involve taxing the first dollar of non-UBI income.⁹

In a UBI scheme where taxes begin at or near the first dollar of private income, most people would enter the no-tax-paying group for no more than a few months at a time, and researchers could not predict in advance who would be most likely to remain in that group longest unless they focused on the disabled—which would defeat the purpose of testing *unconditional* basic income. Therefore, UBI experiments in wealthy nations simply cannot focus on people for whom the difference between gross and net benefit is zero or negligible.

The second option would be to ignore the difference between gross and net benefit even though it is non-negligible. This option can be summarily dismissed because it enormously exaggerates the effects of UBI. The typical net beneficiary in a reasonably-affordable-but-adequate-sized UBI is likely to live in a household that makes substantial private income and benefits by less than half the nominal amount of the UBI, depending on many specific factors about the size and method of financing of the UBI.¹⁰ Ignoring this difference would render any observations of participants behavior almost meaningless as a prediction of what they would do under an actual UBI system. Furthermore, the *rate* at which participants' net benefit decreases as they make more money (or do other things that might increase their tax burden under various possible financing regimes) is likely to have an important effect on their decision-making and their behavior. It simply can't be ignored if the results of the test are going to be at all useful in estimating the effects of a real UBI.

Therefore, any reasonable UBI experiment has to focus on the net and not simply the gross benefit. But as mentioned above, researchers can't levy taxes.

The third option is to simulate the new taxes by reducing participants' grant as their income goes up. But as chapter 3 mentioned, a grant that goes down as income goes up is not a UBI; it's an NIT.

An NIT scheme can create the same after-tax distribution of income as a UBI scheme that happens to have the same marginal income tax rate, and so it is reasonable to say that NIT is a good proxy for UBI in an experiment. But, as Chapter 2 explained, the NIT works differently in some important ways. For example, if a person suddenly finds themselves without income, they will have to inform a government authority that they should start receiving the NIT rather than having a UBI automatically deposited into their account every week—regardless of how much they are receiving in other income and paying in taxes. The practical importance of these differences is a controversial question among people who study or advocate for various forms of BIG. It is the kind of difference we would ideally like to test in an experiment. Instead, experiments will have to assume that these differences are small to use an NIT as an approximation of UBI.

One problem with using NIT to approximate UBI is that it effectively forces the experiment to focus on an at least partially income-tax financed UBI program. In the 1970s, 80s, and 90s, the BIG discussion focused overwhelmingly on an income-tax financed version, often in the form of a flat income tax combined with a UBI.¹¹ But this financing model is not as central to the debate anymore. Many recent proposals focus on rent and resource taxes, banking reforms, wealth taxes, and so on. Many such taxes do not fall directly on net beneficiaries of UBI, but might or might not be passed onto them through the market—once again the kind of thing we would like to test in an experiment rather than to impose on an experiment by assumption.

However, the flat income tax in an experiment has a lot of advantages. It makes the math extremely easy, and whatever type of tax is used, the amount of taxes people end up paying is likely to be heavily correlated with income, so an experiment can use the flat tax as an

approximation for any other tax hopefully without too much loss of generality. Researchers can test a variety of tax rates to account for any uncertainty about exactly what the net tax burden will be under various schemes.

UBI experiments will also be forced to take on the second characteristic of NIT: they will have to give the grant on a household basis rather than an individual basis. Researchers can't simply select a group of individuals at random and give them each a UBI, because most of those individuals live in households, and the effect of UBI on one person in a household where everybody gets a UBI is very different from the effect of a UBI on one person in a household where no one else gets a UBI. Therefore, RCTs will have to draw households at random rather than individuals at random, and they will have to treat those households as a unit.

Furthermore, researchers will have to simulate the taxes needed to support UBI at the household level, because most people pay taxes as households. This will involve reducing everyone's UBI to simulate the increase in taxes as one member's income goes up, effectively treating the UBI as a grant to households rather than to individuals. For example. Imagine a household where only the father receives a private income. A UBI gives a separate income to father, mother, and child, while all of the family's income taxes come out of the father's income. Suppose the father's income rises. Under a fully implement UBI system, everyone's separate UBI grant stays the same, while the father pays more taxes. Under the experimental NIT system, the one NIT grant check they receive as a household unit goes down to simulate the new taxes on the father's larger income. The overall effect on the household's income as a whole is exactly the same. Does this mean that they react the same? We don't actually know. It depends on whether receiving separate UBI's affects the distribution of spending within the household. This is the sort of question we'd like an experiment to answer, but because we are forced to use an NIT as a proxy for UBI, researchers will have to assume that the family will react exactly the same whether the grant is individual- or household-based.

Using NIT to approximate the effect of taxes in a UBI experiment has one obvious further drawback: it stops working once the household income reaches the break-even point. This technique does not help study the effects on net contributors—people who pay more taxes than they receive in UBI and are therefore financially worse off because of it. No one would volunteer for a trial that substantially reduced their income, and a government that forced people to participate in such an experiment would violate the principle of equal protection.

It is reasonable to want to study the effects of a policy in isolation from the effects of the taxes required to finance it, but for an expensive policy like UBI, it's not easy to do so without causing potentially substantial problems.

UBI is an enormous policy. A UBI in the United States, set barely at the poverty-line, would cost nearly 3% of GDP (in net terms), and about 49% of the population would be net recipients.¹² Food Stamps, by contrast, cost only 0.4% of GDP and benefit only 14% of the population in any given year. UBI is not only much larger, but also the net cost has to be borne by a smaller group of people. The effect of this lost income on overall economic activity is not as safely ignored.

The inability to study the effects on net contributors causes at least four problems for a UBI trial.

First, people will move back-and-forth from being net recipients to being net contributors as their income crosses the break-even point. Leaving out the additional taxes that they pay once they become net recipients biases the study to assuming that the financial incentive to earn more private income is larger than it actually would be under a real UBI. There is a good chance that the

marginal effect of these taxes will be small enough to ignore, but once again, that is something we would ideally like to learn from an experiment rather than impose on the experiment as an assumption.

Second, net beneficiaries interact in the market and elsewhere with net contributors. The interactive effects could be substantial because assuming balanced-budget financing, exactly as much money comes out of the economy from net contributors as goes into it via net beneficiaries. The interaction between them is one of those feedback effects that RCTs are completely incapable of examining. The same amount of money is likely to have a smaller effect on the behavior of net contributors than of net recipients, because net contributors have substantially more money. Researchers can use data from other sources to estimate the likely effects on net contributors. There is a wealth of data on how taxation affects behavior. Researchers can then use computer simulations to estimate the feedback effects. Not much of the literature on 1970s NIT experiments involved these kinds of simulations.¹³ And once again, these assumptions of the simulation are things we would ideally like to learn from an experiment rather than impose on it.

Third, this community effect is one that remains unobserved in studies in less wealthy countries as it does studies in wealthier countries, and it affects saturation studies as much as RCTs, causing a particular bias for saturation studies in wealthy countries. In a wealthy country, most communities one might pick for a saturation study will have substantial numbers of both net contributors and net beneficiaries. Researchers will have to counterfactually assume that net contributors pay no new taxes to support the UBI program. Therefore, the community effects will be biased, reflecting the larger budgets of net recipients but ignoring the smaller budgets of net contributors. This imbalance is likely to exaggerate economic activity in the community and therefore exaggerate the opportunities available to net recipients. Again, the effect maybe small, but it is another assumption to impose on the experiment—another caveat to explain.

Fourth, in practical terms, the largest problem with the inability to include net contributors might not be one of biasing the results, but one of helping nonspecialists understand the meaning of the results. Researchers conducting RCTs usually deal with it the inability to study the effects of net recipients in part by confining their sample to people who are very likely to be net recipients—sometimes to people at the very bottom of the net recipient range. They will report results for average comparisons between the control and experimental groups drawn from that subset of the population, but citizens and policymakers will be most interested in how the UBI affects the average person nationwide. If they interpret the numbers they read as being representative of the whole of the population their understanding will be highly biased toward exaggerating UBI's effects for good or bad—even if the study was a completely unbiased estimate of the segment of the population it sampled.

Chapter 6 below shows how this misreading bias had an important effect on the reporting of the labor response in the 1970s experiments. Attention focused on the average difference between control and experimental groups across the various studies (5% to 7.9% decline in labor hours among primary breadwinners) with little or no public discussion that these figures estimated the response of a small subset of the population rather than the response of the entire population. Probably very few nonspecialist readers understood the difference and took it into account. If they did not understand this difference, they had a very biased understanding of the reported results. One computer simulation estimated that the national work-reduction response would be only about one-third of the reduction in the Gary experiment (1.6% rather than 4.5%).¹⁴ I wonder whether numbers like 1.6%—more easily perceived as negligible—would have had a different effect on the discussion of the results at the time.

Researchers conducting and writing about the new round of studies should try not to let that misunderstanding happen again. But it will not be easy to eliminate it. Researchers can include caveats and run simulations, but they can't force the audience to pay attention.

Chapter 6: BIG experiments of the 1970s and the public reaction to them

Between 1968 and 1980, the U.S. and Canadian Government conducted five NIT experiments. They were the world's first major social science experiments. They provide not only inspiration, precedent, and important lessons for the current experiments, but also relevant data, some of which is summarized below. Lessons from these experiments are discussed throughout this book.

Labor market effects of the NIT experiments of the 1970s

Unfortunately, most of the attention of the 70s experiments was directed not at the effects of the policy (how much does it improve the welfare of low-income people) but to one potential side effect (how does it affect labor hours of test subjects). And so that issue takes up most of the discussion here. This section draws heavily on an earlier work, entitled, "A Failure to Communicate: What (if anything) Can We Learn from the Negative Income Tax Experiments."¹⁵

Table 1 summarizes the basic facts of the five NIT experiments. The first, the New Jersey Graduated Work Incentive Experiment (sometimes called the New Jersey-Pennsylvania Negative Income Tax Experiment or simply the New Jersey Experiment), was conducted from 1968 to 1972. The treatment group originally consisted of 1,216 people and dwindled to 983 (due to dropouts) by the conclusion of the experiment. Treatment group recipients received a guaranteed income for three years.

The Rural Income Maintenance Experiment (RIME) was conducted in rural parts of Iowa and North Carolina from 1970 to 1972. It began with 809 people and finished with 729.

The largest NIT experiment was the Seattle/Denver Income Maintenance Experiment (SIME/DIME), which had an experimental group of about 4,800 people in the Seattle and Denver metropolitan areas. The sample included families with at least one dependent and incomes below \$11,000 for single-parent families or below \$13,000 for two-parent families. The experiment began in 1970 and was originally planned to be completed within six years. Later, researchers obtained approval to extend the experiment for 20 years for a small group of subjects. This would have extended the project into the early 1990s, but it was eventually cancelled in 1980, so that a few subjects had a guaranteed income for about nine years, during part of which time they were led to believe they would receive it for 20 years.

The Gary Income Maintenance Experiment was conducted between 1971 and 1974. Subjects were mostly black, single-parent families living in Gary, Indiana. The experimental group received a guaranteed income for three years. It began with a sample size of 1,799 families, which (due to a large drop-out rate) fell to 967 by the end of the experiment.

The Canadian government initiated the Manitoba Basic Annual Income Experiment (Mincome) in 1975 after most of the U.S. experiments were winding down. The sample included 1,300 urban and rural families in Winnipeg and Dauphin, Manitoba with incomes below C\$13,000

per year. By the time the data collection was completed in 1978, interest in the guaranteed income was seriously on the wane and the Canadian government cancelled the project before the data was analyzed.

Table 1: Summary of the Negative Income Tax Experiments in the U.S. & Canada

Name	Location(s)	Data collection	Sample size: Initial (final)	Sample Characteristics	G*	t**
The New Jersey Graduated Work Incentive Experiment (NJ)	New Jersey & Pennsylvania	1968-1972	1,216 (983)	Black, white, and Latino, 2-parent families in urban areas with a male head aged 18-58 and income below 150% of the poverty line.	0.5 0.75 1.00 1.25	0.3 0.5 0.7
The Rural Income-Maintenance Experiment (RIME)	Iowa & North Carolina	1970-1972	809 (729)	Both 2-parent families and female-headed households in rural areas with income below 150% of poverty line.	0.5 0.75 1.00	0.3 0.5 0.7
The Seattle/Denver Income-Maintenance Experiments (SIME/DIME)	Seattle & Denver	1970-1976, (some to 1980)	4,800	Black, white, and Latino families with at least one dependent and incomes below \$11,00 for single parents, \$13,000 for two parent families.	0.75, 1.26, 1.48	0.5 0.7, 0.7-.025y, 08-.025y
The Gary, Indiana Experiment (Gary)	Gary, Indiana	1971-1974	1,799 (967)	Black households, primarily female-headed, head 18-58, income below 240% of poverty line.	0.75 1.0	0.4 0.6
The Manitoba Basic Annual Income Experiment (Mincome)	Winnipeg and Dauphin, Manitoba	1975-1978	1,300	Families with, head younger than 58 and income below \$13,000 for a family of four.	C\$3,800 C\$4,800 C\$5,800	0.35 0.5 0.75

* G = the Guarantee level.

** t = the marginal tax rate

Source: Reproduced from Widerquist (2005)

Scholarly and popular media articles on the NIT experiments focused, more than anything else, on the NIT's "work-effort response"—the comparison of how much the experimental group worked relative to the control group. Table 2 summarizes the findings of several of the studies on the work-effort response to the NIT experiments, showing the difference in hours (the "work reduction") by the experimental group relative to the control group in foregone hours per year and in percentage terms. Results are reported for three categories of workers, husbands, wives, and "single female heads" (SFH), which meant single mothers. The relative work reduction varied substantially across the five experiments from 0.5% to 9.0% for husbands, which means that the experimental group worked less than the control group by about ½ hour to 4 hours per week, 20

to 130 hours per year, or 1 to 4 fulltime weeks per year. Three studies averaged the results from the four U.S. experiments and found relative work reduction effects in the range of 5% to 7.9%.¹⁶

The response of wives and single mothers was somewhat larger in terms of hours, and substantially larger in percentage terms because they tended to work fewer hours, to begin with. Wives reduced their work effort by 0% to 27% and single mothers reduced their work effort by 15% to 30%. These percentages correspond to reductions of about 0 to 166 hours per year. The labor market response of wives had a much larger range than the other two groups, but this was usually attributed to the peculiarities of the labor markets in Gary and Winnipeg where particularly small responses were found.

Table 2: Summary of findings of work reduction effect

Study	Data Source	Work reduction* in hours per year ** and percent			Comments and Caveats
		Husbands	Wives	SFH	
Robins (1985)	4 U.S.	-89 -5%	-117 -21.1%	-123 -13.2%	Study of studies that does not assess the methodology of the studies but simply combines their estimates. Finds large consistency throughout, and “In no case is there evidence of a massive withdrawal from the labor force.” No assessment of whether the work response is large or small or its effect on cost. Estimates apply to a poverty-line guarantee rate with a marginal tax rate of 50%.
Burtless (1986)	4 U.S.	-119 -7%	-93 -17%	-79 -7%	Average of results of the four US experiments weighted by sample size, except for the SFH estimates, which are a weighted average of the SIME/DIME and Gary results only.
Keeley (1981)	4 U.S.	-7.9%			A simple average of the estimates of 16 studies of the four U.S. experiments
Robins and West (1980a)	SIME/DIME	-128.9 -7%	-165.9 -25%	-147.1 -15%	Estimates “labor supply effects.” It goes without saying that this is different from “labor market effects.”
Robins and West (1980b)	SIME/DIME	-9%	-20%	-25%	Recipients take 2.4 years to fully adjust their behavior to the new program.
Cain et al (1974)	NJ	-	-50 -20%	-	Includes caveats about the limited duration of the test and the representativeness of the sample. Notes that the evidence shows a smaller effect than nonexperimental studies.
Watts et al (1974)	NJ	-1.4% to -6.6%	-	-	Depending on size of G and t
Rees and Watts (1976)	NJ	-1.5 hpw** -0.5%	-0.61%	-	Found anomalous positive effect on hours and earnings of blacks.
Ashenfelter (1978)	RIME	-8%	-27%	-	“There must be serious doubt about the implications of the experimental results for the adoption of any permanent negative income tax program.”
Moffitt (1979a)	Gary	-3% to -6%	0%	-26% to - 30%	No caveat about missing demand, but careful not to imply the results mean more than they do.
Hum and Simpson (1993a)	Mincome	-17 -1%	-15 -3%	-133 -17%	Smaller response to the Canadian experiment was not surprising because of the make-up of the sample and the treatments offered.

* The negative signs indicate that the change in work effort is a reduction

** Hours per year except where indicated “hpw,” hours per week.

NJ = New Jersey Graduated Work Incentive Experiment

SIME/DIME = Seattle / Denver Income Maintenance Experiment

Gary = Gary Income Maintenance Experiment

RIME = Rural Income Maintenance Experiment

Mincome = Manitoba Income Maintenance Experiment

SFH = Single Female “head of household.”

Source: Reproduced from Widerquist (2005)

All or most of the figures reported above are raw comparisons between the control and experimental groups: they are not predictions of how labor market participation is likely to change in response to an NIT or UBI. As the second part of this book will explain, there are many reasons

why these figures can't be taken as predictions of responses to a national program. I'll discuss four of them now.

First, the study participants were drawn only from a small segment of the population: people with incomes near the poverty line, about the point at which people are most likely to work less in response to an income guarantee because the potential grant is high relative to their earned income. Thus, the response of this group is likely to be much larger than the response of the entire workforce to a national program. One study using computer simulations estimated that the work reduction in response to a national program would be only about one-third of reduction in the Gary experiment (1.6% rather than 4.5%).¹⁷ Although simulations are an important way to experimental data to what we really want to know, they have important limitations we discuss below.

Second, the figures do not include any demand response, which economic theory predicts would lead to higher wages and a partial reversal of the work-reduction effect. As average labor hours decline, firms respond by bidding up wages, and workers respond by increasing average labor hours. One study using simulation techniques to estimate the demand response found it to be small.¹⁸ Another found, "Reduction in labor supply produced by these programs does tend to raise low-skill wages, and this improves transfer efficiency."¹⁹ That is, it increases the benefit to recipients from each dollar of public spending.

Third, the figures were reported in average hours per week and very often misinterpreted to imply that 5% to 7.9% of primary breadwinners dropped out of the labor force. The reduction in labor hours was not primarily caused by workers reducing their hours of work each week (as few workers are able to do even if they want to). Moreover, few if any workers simply dropped out of the labor force for the duration of the study, as knee-jerk reactions to guaranteed income proposals often assume.²⁰ Instead, it was mainly caused by workers taking longer to find their next job if and when they became nonemployed.

Fourth, the experimental group's "work reduction" was only a relative reduction in comparison to the control group. Although this language is standard for experimental studies, it doesn't imply that receiving the NIT was the major determinate of labor hours. In fact, in some studies, labor hours increased for both groups, and the labor hours of both groups tended to rise and fall together along with the macroeconomic health of the economy—implying that when more or better jobs were available, both groups took them, but when they were less available, the control group searched harder or accepted less attractive jobs.²¹

As we'll see below, most laypeople writing about the NIT experiments assumed any work reduction, no matter how small, to be an extremely negative side effect. But it is not obviously desirable to put unemployed workers in the position where they are desperate to start their next job as soon as possible. It's obviously bad for the workers and families in that position. It's not only difficult to go through but also it reduces their ability to command good wages and better working conditions. Increased periods of nonemployment might have a social benefit if they lead to better matches between workers and firms.

Non-labor-market effects of the NIT experiments

The focus of the 1970s experiments on work effort is in one way surprising, because presumably, the central goals of a UBI involve its effects on poverty and the wellbeing of relatively low-income people, and assessing these issues requires looking at non-labor-market effects.

The experimental results for various quality-of-life indicators were substantial and encouraging. Some studies found significant positive influences in elementary school attendance

rates, teacher ratings, and test scores. Some studies found that children in the experimental group stayed in school significantly longer than children in the control group. Some found an increase in adults going on to continuing education. Some of the experiments found desirable effects on many important quality-of-life indicators, including reduced incidents of low-birth-weight babies, increased food consumption, and increased nutritional content of the diet. Some even found reduced domestic abuse and reduced psychiatric emergencies.²²

Much of the attention to non-labor market effects focused not on the presumed goals of the policy but on another side effect: a controversial finding that the experimental group in SIME-DIME had a higher divorce rate than the control group. Researchers argued forcefully on both sides with no conclusive resolution in the literature. The finding was not replicated by the Manitoba experiment, which found a lower divorce rate in the experimental group. The higher divorce rate in some studies examining SIME-DIME was widely presented as a negative effect, even though the only explanation for it that researchers on either side were able to come up with was that the NIT must have relieved women from financial dependence on husbands.²³ It is at the very least questionable to label one spouse staying with another solely because of financial dependence as a “good” thing.

An overall assessment?

Most of the researchers involved considered the results extremely promising overall. Comparisons of the control and experimental group indicated that the NIT was capable of significantly reducing the material effects of poverty, and the relative reductions in labor effort were probably within the affordable range and almost certainly within the sustainable range.

But experiments of this type were not capable of producing a bottom line. Nonspecialists examining these results might find themselves asking: What *was* the cost exactly? How much were the material effects of poverty reduced? What is the verdict from an overall comparison of costs and benefits?

As Part Two of this book explains in much more detail, experiments cannot produce an answer to these questions. Doing so would involve taking positions on controversial normative issues, combining the experimental results with a great deal of nonexperimental data, and plugging it into a computer model estimating the micro- and macroeconomic effects of a national policy. The results of that effort would be driven more by those normative positions, nonexperimental data, and modeling assumptions than by the experimental results that such a report would be designed to illustrate.

A qualitative grasp of the complexity of the results and what they are likely to indicate about a national policy is about the best understanding a researcher can expect from an audience of nonspecialists. Communicating such an understanding is no easy task—as the public reaction the NIT experiments reveals.

Public reaction to the release of NIT experimental findings in the 1970s

As promising as the results were to the researchers involved the NIT experiments, they were seriously misunderstood in the public discussion at the time. But the discussion in Congress and in the popular media displayed little understanding of the complexity. The results were spun or misunderstood and used in simplistic arguments to reject NIT or any form of guaranteed income offhand.

The experiments were of most interest to Congress and the media during the period from 1970 to 1972, when President Nixon's Family Assistance Plan (FAP), which had some elements of an NIT, was under debate in Congress. None of the experiments were ready to release final reports at the time. Congress insisted researchers produce some kind of preliminary report, and then members of Congress criticized the report for being "premature," which was just what the researchers had initially warned.²⁴

Results of the fourth and largest experiment, SIME/DIME, were released while Congress was debating a policy proposed by President Carter, which had already moved quite a way from the NIT model. Dozens of technical reports with large amounts of data were simplified down to two statements: It decreased work effort and it supposedly increased divorce. The smallness of the work disincentive effect hardly drew any attention. Although researchers going into the experiments agreed that there would be some work disincentive effect and were pleased to find it was small enough to make the program affordable, many members of Congress and popular media commentators acted as if the mere existence of a work disincentive effect was enough to disqualify the program. The public discussion displayed little, if any, understanding that the 5%-to-7.9% difference between the control and experimental groups is not a prediction of the national response. Nonacademic articles reviewed by one of the authors²⁵ showed little or no understanding that the response was expected to be much smaller as a percentage of the entire population, that it could potentially be counteracted by the availability of good jobs, or that it could be the first step necessary for workers to command higher wages and better working conditions.

The United Press International simply got the facts wrong, saying that the SIME/DIME study showed that "adults might abandon efforts to find work." The UPI apparently did not understand the difference between increasing search time and completely abandoning the labor market. The Rocky Mountain News claimed that the NIT "saps the recipients' desire to work." The *Seattle Times* presented a relatively well-rounded understanding of the results, but despite this, simply concluded that the existence of a decline in work effort was enough to "cast doubt" on the plan. Others went even farther, saying that the existence of a work disincentive effect was enough to declare the experiments a failure. Headlines such as "Income Plan Linked to Less Work" and "Guaranteed Income Against Work Ethic" appeared in newspapers following the hearings. Only a few exceptions such as Carl Rowan for the *Washington Star* (1978) considered that it might be acceptable for people working in bad jobs to work less, but he could not figure out why the government would spend so much money to find out whether people work less when you pay them to stay home.²⁶

Senator Daniel Patrick Moynihan, who was one of the few social scientists in the Senate, wrote, "But were we wrong about a guaranteed income! Seemingly it is calamitous. It increases family dissolution by some 70 percent, decreases work, etc. Such is now the state of the science, and it seems to me we are honor bound to abide by it for the moment." Senator Bill Armstrong of Colorado, mentioning *only the existence* of a work-disincentive effect, declared the NIT, "An acknowledged failure," writing, "Let's admit it, learn from it, and move on."²⁷

Robert Spiegelman, one of the directors of SIME/DIME, defended the experiments, writing that they provided much-needed cost estimates that demonstrated the feasibility of the NIT. He said that the decline in work effort was not dramatic, and could not understand why so many commentators drew such different conclusions than the experimenters. Gary Burtless (1986) remarked, "Policymakers and policy analysts ... seem far more impressed by our certainty that the efficiency price of redistribution is positive than they are by the equally persuasive evidence that the price is small."²⁸

This public discussion certainly displayed “a failure to communicate.” The experiments produced a great deal of useful evidence, but for by-far the greatest part, it failed to raise the level of debate either in Congress or in public forums. The literature review reveals neither supporter nor opponents who appeared to have a better understanding of the likely effects of the NIT and UBI in the discussions following the release of the results of the experiments in the 1970s.²⁹

Whatever the causes for it, an environment with a low understanding of complexity is highly vulnerable to spin with simplistic if nearly vacuous interpretation. All sides spin, but in the late 1970s NIT debate, only one side showed up. The guaranteed income movement that had been so active in the United States at the beginning of the decade had declined to the point that it was able to provide little or no counter-spin to the enormously negative discussion of the experimental results in the popular media.

Whether the low information content of the discussion in the media resulted more from spin, sensationalism, or honest misunderstanding is hard to determine. But whatever the reasons, the low-information discussion of the experimental results put the NIT (and, in hindsight, UBI by proxy) in an extremely unfavorable light, when the scientific results were mixed-to-favorable.

The scientists who presented the data are not entirely to blame for this misunderstanding. Neither can all of it be blamed on spin, sound bites, sensationalism, conscious desire to make an oversimplified judgment, or the failure of reports to do their homework. Nor can all of it be blamed on the people involved in political debates not paying sufficient attention. It is inherently easier to understand an oversimplification than it is to understand the genuine complexity that scientific research usually involves no matter how painstakingly it is presented. It may be impossible to communicate the complexities to most nonspecialists readers in the time a reasonable person to devote to the issue.

Nevertheless, everyone involved has a responsibility to try to do better next time.

The rest of this book is an effort to help reduce similar misunderstandings in future experiments. It is aimed at a wide audience because it focuses the problem of communication from specialists to nonspecialists. We hope to help researchers involved in current and future experiments design and report their findings in ways that are more likely to raise the level of debate; to help researchers not involved in the experiments raise the level of discussion when they write about the findings of the experiments, to help journalists understand and report experimental findings more accurately; and to help interested citizens of all political predispositions see beyond any possible spin and media misinterpretations to the complexities of the results of this next round of experiments—whatever they turn out to be.

Chapter 7: New experimental findings 2009-2013

The last of the 1970s NIT experiments came to an end in 1980. Academics continued to discuss the results for more than a decade, but no new BIG experiments were conducted, and no new results were reported until the late 2000s and early 2010s when two new experiments were conducted just as new results from one of the 1970s experiments were finally released.

Canada’s Mincome experiment was cancelled before most its findings were assessed. As many as 1,800 boxes of file folders were left in examined until 2009 when a researcher named Evelyn Forget got a grant to begin reopening them. Perhaps, she did a better job of explaining the findings in a way that people understood, or perhaps the political situation at the time made for a more receptive audience. But whatever the reason, the newly released Mincome findings had a much more positive impact on the UBI debate than NIT experimental findings released in the

1970s. Forget dubbed Mincome's saturation site (Dauphin, Manitoba) "the Town With No Poverty," and the media picked up on it. Media reports stressed the effects (rather than the side effects) of Mincome. These effects included reductions in hospitalizations, especially for mental health and accidents. Forget estimated the national savings that would occur if the decline in hospital visits was replicated nationally.³⁰ Media reports discussing the labor market impact did so in context, even discussing how the lack of pressure to find another job helped people land the right job.

The first UBI experiments of the twenty-first century were conducted in Namibia (2008-2009) and India (2011-2013). They differed from the 1970s experiments in three important ways. First, they focused on UBI rather than NIT, reflecting the change in the discussion of BIG over the intervening 30 years. Second, they were funded primarily by private institutions rather than the government. Third, both of them took place at a time when BIG was not a major part of the political discussion in the countries where they were conducted.

The Namibian study found extremely promising results. Results included significant decreases in household poverty, child malnutrition, underweight children, household debt, crime, and so on. Results also included significant increases in economic activity, access to medication and healthcare, school attendance, and household savings. Predicted effects of increased alcohol consumption did not come true. Probably the most striking difference between the Namibia project and the NIT experiments was that the work-effort response was positive. That is, people receiving UBI worked more.³¹ The expected explanation was that the depth of poverty and the level of unemployment in Namibia make it hard for people to work as much as they might want to. With more of their basic needs met and more economic activity in the area, people were able to work more.

The Indian project found similar promising results. Results included significant decreases in illness, child labor, household indebtedness, and so on. Results also included significant improvements in food consumption, medical treatment, school attendance, school performance, household savings, and so on. Like the Namibia study, the India study found that people receiving UBI worked more than people in the control group. They also invested more in self-employment activities.³²

The twenty-first-century reports from Mincome and the reports from India and Namibia were well reported and better understood in the press. All three seemed to increase political attention to UBI in the countries where the studies were conducted and around the world.

Part Two: The place of experiments in the political economy of UBI

Chapter 8: Why UBI experiments cannot resolve much of the public disagreement about UBI

An article in the December 2016 issue of *the MIT Technology Review* had the wildly inaccurate headline “In 2017, We Will Find Out If a Basic Income Makes Sense.”³³ Its most laughable inaccuracy was that experiments were barely getting underway in 2017. None of the experiments the article mentioned (nor any others) had plans to release any findings at all in 2017 (nor did they). Unfortunately, the headline included a much more important inaccuracy: the naïve belief that UBI experiments are capable determining whether UBI “makes sense.” Also, unfortunately, the headline’s overblown expectation is representative of a lot of reporting as the experiments get underway.

It is understandable that citizens and policymakers *want* answers to the big questions, but it’s an enormous problem if they actually *expect* UBI experiments to produce answers to these big questions and worse still if answers are misunderstood as showing more about the big questions than they actually do. Overblown expectations are a primary reason why experimental results are vulnerable to misunderstanding, sensationalism, and spin. Nonspecialists need to understand the difference between what experiments can do and any attempt to “Find out if a Basic Income makes sense,” if they are going to understand experiments at all.

The belief that a UBI experiment can provide a definitive answer to the question of whether to introduce UBI rest on three false presumptions: 1. People disagree about UBI primarily because they disagree about what its effects might be. 2. These disagreements about effects stem from a lack of available evidence. 3. An experiment will provide that missing evidence.

This chapter dispels those false presumptions as it gives three arguments. First, the public disagreement about UBI is more of an ethical debate about the desirability of its effects than an empirical debate about what its effects are. Second, disagreements about UBI’s effects don’t stem primarily from a lack of available evidence. A lot of evidence already exists. Third, experiments cannot provide the most important missing evidence. They will only add a small amount to the existing body of evidence and leave many of the most important remaining empirical questions about UBI unanswered.

Thus, this chapter explores the difference between the questions that need to be resolved to make a decision for or against UBI and the answers UBI experiments can provide. Exploration of this disconnection might appear negative, but it’s necessary to understand how to create the best experiment and get the most out of it.

Experiments cannot resolve the basic disagreement about UBI because it is largely an *ethical* rather than an *empirical* disagreement. That is, ethical disagreements about the desirability of UBI’s effects play a much larger role in the UBI debate than empirical disagreements about what UBI’s effects are. This is not because people don’t care about evidence, but as argued below, because UBI’s likely effects are well-enough understood and the moral desirability of those effects is controversial enough to make the ethical part of the argument pivotal.

Experiments are empirical studies. They can provide evidence to help answer empirical questions like *what does this do?* But they cannot provide the answer to ethical and subjective questions, such as *do we want what this does?* And the UBI debate turns largely on whether people of different ethical perspectives like what it does.

For example, UBI supporters tend to believe either that it is good for everyone to be free from the threat of poverty including non-wealthy people who might refuse to take jobs or that the possibility that non-wealthy people who might refuse to take jobs is not bad enough to compel sacrificing other goals that UBI might achieve. Opponents tend to believe it is wrong for anyone (who isn't independently wealthy) to get anything without taking a job. These positions differ on basic ethical premises—as do position in many similar disagreements over UBI. No empirical study of the practical effects of a UBI will determine whether these two incompatible ethical beliefs are right or wrong.

There are other reasons both ethical and empirical to support or oppose UBI. Part Three has an extensive discussion of claims made by both sides. Any of those claims could be the basis of support or opposition. But some element of this kind of ethical divide exists in virtually all UBI debates. This kind of ethical divide exists in the background of most discussions over UBI's ability to achieve almost any goal, and people who haven't made up their minds on UBI often bring up concerns that are closely related to this and similar ethical disputes.

Empirical research can find evidence that is useful to people debating ethical issues. For example, if research was to find out that the average person spends just as much time on a job with UBI as without, at least some people who oppose it because it allows non-wealthy people to live without taking a job would probably change their minds. But not all people who hold that position will be swayed. Some might believe non-wealthy people need to work more than they are working now. Others might oppose UBI because they think it is wrong to allow even the possibility that a non-wealthy person might refuse a job offer.

Similarly, if empirical research found that a given level of UBI caused a decrease in employment so large that it threatened UBI's sustainability, any UBI supporters who aren't extremely short-sighted would drop their support for UBI or at least for that level of UBI if they were unable to suggest a policy to counteract that unsustainability.

Yet, experiments in wealthy countries are unlikely to show either result. Past evidence strongly indicates that low-wage workers in wealthy countries will spend less time in employment but not so much less that UBI will become unsustainable. If experiments are consistent with a decline in labor hours in that range, supporters are likely to say UBI has passed the test, and opponents are likely to say that it has failed. People whose opinions are in the middle might be more open to changing their minds or making up their minds for the first time based on more where in that range the estimate falls. But subtle results are not likely to be a huge deviation from what we can already estimate from existing evidence. And it's possible that responses of people in the middle will be affected less by the small amount such results would add to past evidence than by who wins the spin wars that are likely to follow the release of experimental findings.

If we want a UBI experiment, we need to accept that it will not settle the major ethical divides between supporters and opponents, nor will it prove either of their positions untenable.

It will also have to deal with the problem of separating empirical from ethical claims. It's not easy to evaluate whether UBI works with making ethical judgments about how to evaluate performance. I'll put this issue away and return to it in Chapter 12.

An enormous amount of evidence about UBI's cost-effectiveness already exists. A great deal of research about UBI or NIT has been done. This research includes thousands of books and articles on various aspects of UBI's effects as well as seven large-scale trials conducted worldwide between 1926 and 2013. In addition, studies of policies of varying degrees of similarity, such as the Alaska Dividend, conditional cash transfers, citizens pensions, tax credits, and many others provide information that can be used to estimate UBI's effects.³⁴

My impression—after studying UBI for more than 20 years—is that the better one grasps existing evidence, the more likely one’s decision comes down to ethical issues. I can say that the right UBI scheme will be sustainable and will do things people with an ethical position similar to mine want it to do, but it will also do things that people with different ethical positions do not want it to do.

There are also reasonable people in the middle who might well be swayed by new evidence, but many people in the middle aren’t familiar with the existing evidence, and it is uncertain that a new experiment will provide the most important piece of missing evidence they have been looking for, as Part Three will argue.

Existing evidence is not assembled in any one spot nor is most of it easily accessible to nonspecialists. The fullest summaries of existing evidence are in books written by specialist supporters, such as Malcolm Torry, Annie Miller, Philippe Van Parijs, Yannick Vanderborght, Guy Standing, and many others.³⁵ Of course, books by supporters might be subject to confirmation bias. Although there are many well-researched criticisms of UBI, I don’t know of any book-length examination of existing empirical evidence by an opponent, and so interested nonspecialists will have to rely on supporters to provide the most extensive treatments of the evidence.

Despite the enormous amount of evidence available in the relevant social science literature and the availability of good summaries, a substantial part of the current discussion of UBI among citizens and policymakers still goes on in ignorance of existing evidence. In fact, a lot of clearly false claims easily contradicted by evidence are regularly repeated in the debate. For example, many people continue to claim that a poverty-level UBI would cost 15-20% of GDP, when the actual amount is estimated to be about one-sixth of that figure, less than 3% of GDP.³⁶

It’s possible that a better way to raise the level of discussion would be to find a way to more widely inform people about existing evidence. Of course, it would be hard for any government to attempt to provide a *definitive* treatment of existing evidence. Any effort is likely to be criticized as biased from one side or the other or both.

Important gaps in the existing evidence do remain. Experiments can help fill in some of those gaps, but as Part Three shows, experiments are only capable of testing a small subset of what we really want to know about UBI. They add a small (but valuable) amount of information to the existing body of evidence on UBI’s probable effects. And many of the biggest and most important gaps in the existing evidence are not things that UBI experiments are capable of addressing. Neither these gaps nor the potential for UBI experiments to fill them in are well-understood by nonspecialists including some of the reporters currently writing about the experiments.

In addition, as Part Three discusses, the testable aspects of UBI are often—if not always—only testable in an indirect, partial, or statistically biased manner. Experimental findings have to be combined with other evidence to provide answers to relevant questions in the debate over this policy. That combination of experimental and nonexperimental evidence is not usually the responsibility of the researchers who conduct the experiments, but later chapters will argue that it might be necessary to reduce the risk that citizens and policymakers will not understand what the experiments imply about what they really want to know.

The decision to conduct a UBI experiment should be made with full knowledge of all these limitations. If we want a UBI experiment, we need to accept not only that it is incapable of settling the major ethical divides between supporters and opponents, not only that it is highly unlikely to prove either of their positions untenable, but also that it is unlikely to provide a large enough addition to existing evidence to give a compelling reason for massive numbers of people in the middle to shift their opinions significantly. Experiments will make a small addition to existing

evidence about what UBI can do. If the results are well-communicated and well-understood, their best realistic hope is to raise the level of discussion among people on all sides in the current debate by increasing both the evidence available to them and their understanding of it. This is a good reason to do an experiment, but it is far short of the definitive test people have reason to want and some seem to be expecting.

Specialists would be mistaken to think everything I've said in this chapter is too obvious to mention. That mistake is a central reason this book is necessary. Citizens and policymakers have to be free of any false hopes for UBI experiments if their decision to conduct one is to be based on what an experiment can actually do.

Chapter 9: The political economy of the decision to have a UBI experiment

I've argued that everyone with an interest in UBI experiments needs to understand the political economy of the public discussion of UBI. This effort begins with an understanding of the strategic decision to have a UBI experiment. And the decision *is* strategic.

There are many scientific reasons for a UBI trial. It can shed some light on at least some of the controversial questions about UBI's practical effects, but their existence is not necessarily the reason trials are happening. There is no movement of people who are curious about UBI, who have questions about its effects, and who would like to examine whatever effects a trial is capable of examining. The demand for trials is a response to the growing UBI movement, which is made up almost entirely of people who are already convinced UBI works and want it introduced. The spread of experiments is a strange victory for that movement.

UBI supporters who favor trials do so for strategic reasons. That is, they hope trials will help build support for UBI and eventually lead to its introduction. To say their reasons are strategic is not to say that UBI supporters want anything less than a scientific study. But the desire for a scientific study does not make the decision to favor a trial any less strategic. The strategic hope is that good, scientific inquiry into the issue will demonstrate the efficacy of the program, build the movement, and lead to its introduction.

Although trials have great promise, they are a risky strategy for the UBI movement. Trials can't address all of the most important questions about UBI. The streetlight effect could deflect attention from more important onto lesser important issues. A lack of understanding of the limits of experiments could be misunderstood with negative consequences. No matter how positive the results might be, they are vulnerable to negative spin by people who reject the very idea of UBI. The version of BIG that is tested is likely to be very different from the version that people in the movement most favor. And the existence of a trial can deflect political momentum.

UBI experiments are part of the political process in part because they are too large to be funded by a simple research grant. Although the three UBI trials conducted or underway so far in less wealthy nations (Kenya, India, and Namibia) have been funded privately, only one experiment in a wealthy nation has been privately funded so far (the Y Combinator experiment in California). However, none of these were primarily financed by scientific research grants from any of the regular grant-giving institutions, but by well-funded institutions that are closely connected to the UBI movement. That is, they are an outcome of the political process.

Most UBI trials are even closer to the political process. For the most part, the decision to have one rests with politicians. All five of the 1970s experiments in the United States and Canada were initiated by acts of law at the national level. The Finnish, Scottish, and Canadian experiments are as well. Other proposals for UBI being floated around the world right now are connected to national legislatures.

Why are so many politicians around the world suddenly so interested in trials? There are a variety of possible reasons, some better and some worse. Consider four of them.

First (and least likely), a politician might support a trial to discredit the movement. Although the results of a trial can be negatively spun and at least at times in the past have had negative effects on the UBI movement, this motivation is extremely unlikely because it's too risky for politicians who oppose UBI. Just by supporting a trial, they risk alienating their UBI-opposing constituents. Just by talking about a trial they bring media attention to a policy they oppose. There is no negative publicity, as the saying goes. And by conducting a trial, they commit years of funds to a strategy that might well backfire on them if they are unable to control how the trial is perceived by the public. Any UBI opponent with the power to use such an elaborate strategy to discredit UBI is probably better off using that power to try to keep UBI out of the mainstream dialogue.

Second, politicians might be driven by pure scientific curiosity. UBI is hotly debated, partly because its effects are controversial. A trial can help resolve some of that controversy and raise the debate while promoting science. This motivation isn't terribly likely either. Probably, most politicians are politicized. If they are going to support a trial, they have some partisan interest in the outcome of a trial or at least an interest in the constituency demanding the trial.

Third, politicians might supporter UBI and believe that a trial will ultimately be good for the movement. If there are enough committed UBI supporters in government to pass a law instituting a trial of UBI, it's worth asking why they don't just skip the trial and pass a law introducing a full UBI right away. The most probable answer is that not enough of their constituents are behind them right now.

UBI is no small idea. Virtually any substantial version of BIG would be an enormous change to any country's public policy system. Despite UBI's growth, the idea is still a minority opinion in most countries. It would be an enormous risk for politicians to make such a change without the confidence that they had a substantial constituency behind them. Politicians might hope that a successful trial can help build that coalition, and so the politicians opting for a trial rather than the immediate introduction of UBI might nevertheless share some of the motivations of UBI supporters.³⁷ If the motivation for UBI experiments is to try to gather political support, then the decision to conduct a trial is definitely a strategic one. These enormous scientific undertakings are being conducted for advertising in the form of demonstrating feasibility.

Fourth, a trial could be some kind of consolation prize for the UBI movement. While the UBI movement wants the support of politicians, politicians want the support of the UBI movement. A consolation prize could be politicians' way of saying that the movement has grown enough to be taken seriously and enough that at least some political parties find it necessary to seek the support of that movement. But the constituency has not grown enough to get UBI introduced. The consolation prize of a UBI experiment may be the next best politically-feasible thing that politicians can do at this point to get the UBI movement to support them.

UBI trials work very well as a consolation prize for at least two reasons. First, they are much less costly in terms of both money and political capital. Second, few other consolation prizes are readily available because a large portion of people in the UBI movement today are unlikely to

be placated by gradual steps in the direction of BIG. Many of them want a big change now and are suspicious that any small change in the existing system is truly a step in the direction of UBI.

Although one would have to be cynical to the point of naiveté to put much stock in the first of these four possible motivations, party support for UBI trials might well be driven by a mix of the second, third, and fourth reasons. It is even possible that politicians themselves don't know the extent to which they are driven by each of those motives. They might hope that the trial will help the UBI movement succeed while in the meantime, they're anxious to secure the movement's support of their party, and they feel good about supporting science.

A danger for the UBI movement comes along with these possible motivations. That is, trials might end up deflecting political momentum away from UBI.

Politicians have a massive incentive to find *the cheapest way to tell you yes*. Even the most well-meaning politicians might feel some of the pressure of the political incentive structure that pushes in this direction. They might *want* to support the UBI movement's cause (full implementation), but they *need* to get the UBI movement to support their cause (reelection). The difference in cost (both monetary and political) between a UBI trial and actual implementation is so enormous that politicians can far more easily deliver a trial and so pressure exists to favor trials. From the politicians' perspective, this is a triple win: they gain a constituency, support scientific research, and take action that might someday lead to the introduction of a policy they sympathize with. Politicians might not be fully aware of the extent to which they are affected by each of these motivations.

Once a trial is in place, it can become a temporary barrier to full implementation. A good trial can last three-to-seven years. Having said yes to a trial, the politician now has the perfect excuse to say no to implementation for that entire period. You got the trial you asked for; it only makes sense to wait for the results before taking the next step. Three-to-seven years is a long time in politics. The movement could peak during that period. Sympathetic parties could lose power.

The NIT experiments of the 1970s might well have been a barrier to the introduction of some form of BIG in the United States when a bill was active in Congress in 1971 and 1972 before any of them were completed. The then-unfinished trials were certainly not the most important barrier to the introduction of BIG at that time, but the issue did not seriously return to Congressional politics in the following years. And when the results finally came out they were negatively spun to discredit BIG, which left the political mainstream for nearly 40 years.³⁸

UBI supporters should not assume that the sad history of the 1970s trials will repeat itself. But it provides a warning. They have to understand the experiments and their place in the UBI debate—as do their opponents and the researchers involved with the experiments.

Chapter 10: The chain of misunderstanding between experimenters and their nonspecialist audience

This chapter attempts to explain why UBI experiments are so vulnerable to misunderstanding, misuse, spin, and the streetlight effect in an effort to help anyone interested in the current round of experiments avoid those problems this time. These vulnerabilities are not any one person or group's fault. They happen because the nonspecialist who create the demand for experiments and the specialists who conduct the experiments have great difficulty understanding each other, and they are separated by a long and difficult chain of connections. Essentially, we're

playing the telephone-game situation, in which each degree of separation adds potential to add misunderstanding, and we're playing it with inherently difficult information. Consider that chain of connections and how it leads to these problems.

I've argued that the demand for experiments begins with citizens' dialogue about UBI. But going back to the first inception of NIT experiments in the late 1960s, that dialogue is seldom driven by people saying things like, "We need an answer to questions x, y, and z about UBI," but by people saying things more like, "We want a trial of UBI," with little discussion of what specifically they want to learn from that trial beyond some vague notions like whether it "makes sense."

The citizens who create the demand for trials are not social science experts. Citizens might not know what questions experiments can and cannot address. I've argued that most citizens are interested in the big questions, an overall verdict on UBI's efficacy. They might need help understanding what pieces of empirical evidence are most important toward those big questions and how any particular piece of evidence relates to those big questions. They will probably count on the researchers conducting the study to decide what questions to address and how to address them, and they might presume or at least hope that these experts will be able to anticipate the questions they want answered and translate that evidence into the right answers.

Politicians, rather than the people most closely involved in the discussion that creates demand for UBI experiments, make the decision to have a trial. Only a few of them will be closely connected to that discussion. They might be interested in a different definition of UBI than the one used by most UBI supporters. Whatever UBI model politicians decide to test, they cannot be counted on to know what questions are most relevant to the citizens involved on any side of the public discussion of UBI. Often, they seem to have no specific questions in mind, and when they do, the questions they ask are not always the same as the questions most important to citizens involved in the public discussion.

For example, U.S. decisions made in the 1960s and 1970s to fund experiments seem to be driven by a desire to have an NIT experiment rather than the desire to answer any particular question about NIT. The Finish experiment was created to focus largely on the question of whether people currently living on social assistance would be more likely to accept employment if they had a UBI instead. This question is relevant, but it's hardly the heart of the UBI debate, and it is very far removed from the big questions like whether UBI is cost-effective overall.

Once politicians make the decision to have an experiment, they hire managers or consultants to appoint a group of social scientists to design and conduct the study. These social scientists are, therefore, separated from the public discussion by several degrees. Each of these degrees of separation has potential to add misunderstanding.

The researchers hired to conduct the trial might or might not be well-versed in the dialogue. They might or might not research it or consult closely with people involved in the discussion. As mentioned above, the vast majority of research specialists are not fools; they will look for evidence that makes a positive contribution to the body of knowledge about UBI. But there are at least four reasons why they might nevertheless focus on different aspects of the issue than the people involved in the debate focus on.

First, social scientists tend to look at research questions very differently than nonspecialists. Nonspecialists tend to want a verdict: you're the expert. Run a test. Tell me if it passes the test. What's the verdict? Up or down? To social scientists, it's obvious that no single study is very likely to produce a decisive verdict on any social science issue. We already have a great body of knowledge about UBI. Let's find out what this experiment can add to that body of

knowledge. Pushing out the body of accumulated knowledge in any way possible is a very important role for scientists. But the question of what we can add to the body of knowledge about this policy is very different than the question of what is the overall, up-or-down verdict on this policy, nor is it easy for a nonspecialist to understand if and when the connection might be a difficult one.

Second, social scientists have no particular expertise in discovering the questions that concern others. Their expertise is in applying the tools they know to questions those tools are most suited to address. Politicians hired them, knowing their area of expertise is to conduct an experiment with many inherent limitations. Social scientists might reasonably assume that they have been hired to do what they do best and to use this experiment to examine the questions it can best address. But, of course, the streetlight effect simply *is* the focus on what researchers do best and/or what an experiment can best observe rather than on what questions most need to be answered.

Third, social science researchers have a strong interest in being seen by their peers as doing something *scientific*. The general climate in most of the social sciences is that quantitative research is somehow more scientific than qualitative research. Studies reporting numbers—the more measurable the better—are seen as more scientific than those reporting less quantifiable observations. In addition, RCTs are seen as being more scientific than saturation studies even if a saturation study produces more relevant results to the issue being studied. Unfortunately, as Part Three discusses, good answers to some of the most important questions in the UBI debate need to be more qualitative than quantitative.

Fourth, specialists—like everyone else including me and you—tend to be biased in favor of believing what they do is important. If so, they are likely to believe that whatever questions their experiment can address are more important than they actually are. They might underemphasize (to themselves and to others) the importance of all those questions that the experiments cannot address or all the differences between experimental findings and its implications about the centrally important questions in the evaluation of UBI as a policy.

This reasoning indicates the possibility that specialists conducting UBI experiments will be most interested in different questions than nonspecialists—the citizens and policymakers involved in the discussion of UBI. This difference in concern would not be crucial, if everyone understood it. Nonspecialists might be disappointed to learn the extent to which, instead of a decisive, overall evaluation of the policy, UBI experiments produced a small improvement in the existing knowledge about a few of the questions relevant to that overall evaluation, but as long as they understand the research and its implications about that overall policy evaluation, it will raise their level of understanding.

Unfortunately, the telephone game begins again as experimental findings make their way into the public discussion.

Researchers usually take other researchers as their primary audience. When they do, they write in the exacting academic terms familiar to other researchers and leave out the background knowledge familiar to other researchers in their respective fields but not necessarily familiar to people outside of their field. As Chapter 1 mentioned, most researchers are professionals at writing for each other and amateurs at writing for nonspecialists. Some are very good at it, but researchers have no special training in communicating with people outside their field, and many excellent researchers are not good at all at communicating with nonspecialists.

The findings of U.S. and Canadian experiments released in the 1970s and early 1980s were mostly released in specialist-to-specialist publications, such as academic monographs and journals,³⁹ which are dense and difficult for nonspecialists.

Hopefully, the studies will produce at least some reports aimed at general audiences. They might even employ science communication specialists to report the results in language that nonspecialists can best understand. However, reports on experimental findings—even if written by science communications specialists—might not focus at all on bridging the gap in understanding that this book is primarily focused on.

Instead, research reports often focus on helping nonspecialists understand the results on their own terms rather than on relating those results to the questions that most concern nonspecialists. That is, reports might help people understand how the behavior of the control group differed from the behavior of the experimental group in the ways that researchers were capable of studying. Reports might not breach the very complex and difficult effort it would take to explain how much (and how little) these differences say about the likely overall effects of a national UBI in the areas of most concern to nonspecialists trying to evaluate UBI as a policy option.

Why not? There are several possible reasons.

First, it's not necessarily their job. Unless specifically instructed, it is not usually the job of researchers or of science communication specialists to find out what questions interest other people. Their job is to conduct research and explain the findings of that research. If our political process hires specialists to do job A, we cannot blame them for neglecting our unspoken need for them to do job B as well.

Second, what is obvious to specialists is not always obvious to nonspecialists. These studies are short-term. They do not capture community effects. They produce indirect and partial inferences about the national application of a policy. They do not address all of the important claims needed to fully evaluate UBI as a policy. From one specialist to another all of this might seem too obvious to mention, or it might seem to merit no more than a dry list of caveats so that other specialists know that the researchers conducting the study were aware of these limitations. If specialists are unaware how poorly nonspecialists understand these issues, they might not even mention them, much less work through all the implications of these limitations drawing on theory and data from other sources to work through these issues to connect experimental results to the questions nonspecialists want answered.

Third, as mentioned above, researchers—like everyone else—are biased in favor of believing that what they do is important. A report emphasizing all the barriers between the experimental results and the things we really want to know would make the experiments look less valuable than they would look in a report that ignored or downplayed those differences. Similarly, a report emphasizing how much theory and data from other sources was necessary to connect the experimental results to the evaluation of the actual effects of a national policy would make the experiments themselves look less valuable.

Fourth, as also mentioned above, social scientists sometimes feel pressure to be seen doing something scientific, which is often conflated with doing something quantitative. The effort to discuss the limitations of experimental findings in order to connect them with answers to the questions nonspecialists most want answered will involve doing more qualitative, nonacademic discussion.

Whether for these reasons or others reports about the U.S. experiments in the 1970s were primarily aimed at other researchers, and they overwhelmingly stressed the differences between

the behavior of the control and experimental groups rather than the part these play in understanding how to evaluate BIG as a potential national policy.⁴⁰

No matter how well-written reports might be, they face the inherent problem that the information they are trying to convey simply is complex and difficult. After all, any nonspecialist who learns what specialists know becomes a specialist. Some amount of the complex implications of a UBI trial simply will be missed by most nonspecialists. The difficult trick is to get them to understand *enough*. That task is not usually impossible, but it isn't ever easy. Weeding through the complexity of the issue to determine what is enough and figuring out how to communicate it is extremely difficult, and it's not necessarily anyone's job.

An oversimplification is inherently easier to understand than true complexity.

People reading about UBI experiments from any source might be biased toward finding oversimplification just because they're looking for something they can understand. They also might be biased in this direction by what we might call "professional deference." By this, I mean the mistaken belief that expert findings are more definitive than even the specialists themselves might believe. In everyday conversations, if one person says several bad things about an idea, they are implying that it's a bad idea. That's not necessarily true in a complex research report. But readers who are most interested in the big question might well make inferences like that. They might take every positive-sounding result as a vote for and every negative-sounding result as a vote against the policy.

Whether or not researchers conducting experiments produce reports attempting to explain that complexity directly to nonspecialists, most nonspecialists (i.e. most citizens and politicians) will get most of their information about the study not from research reports but from journalists, bloggers, and columnists (call them popular writers),⁴¹ creating yet another degree of separation, and one that involves opportunities for spin and sensationalism.

Popular writers might well be professional writers, but few of them are professional social scientists. Only a few of them will have much more expertise than the public they write for. They might struggle to understand research reports even on their own terms. They might be incapable of doing the complex analysis necessary to relate reported differences between control and experimental groups to probable outcomes for a national UBI. That is, they might have some of the same problems as their readers in understanding the results of UBI trials.

If their understanding is oversimplified, they are likely to be biased toward sensationalism. Whether it is out of professional deference, a desire to attract more readers or the inherent difficulty of the material, many recent reports about the UBI experiments getting underway now have been sensational, claiming the experiments will have far more conclusive results than they really do (CITE). The reporting in the 1970s on NIT experiments was overwhelmingly sensational.⁴²

Most likely, some writers, politicians, and even some of the researchers themselves will—consciously or unconsciously—spin the results to the advantage of one side or other in the debate. "Spin"—as I use the term—is not necessarily deceptive. To spin is to present information in a way that favors one or another interpretation of it. A person (like me) who is convinced UBI is a good idea cannot present what they know honestly without also putting UBI in a favorable light. The same is true for opponents. This honest spin is not unethical, but it is a source of misunderstanding as information goes through the telephone game.

Spin becomes dishonest when people knowingly overemphasize one side of the issue over another. This kind of spin can still be unconscious if it stems from a bias toward recognizing favorable evidence as more important than unfavorable evidence, but it is deceptive and can be a big source of misunderstanding.

Some people might not even look for the evidence as a way to improve their understanding but as a source of ammunition to use to defend their perceived position on the issue.

Most citizens will get their information from popular writers with all of this potential for misunderstanding, oversimplification, sensationalism, and spin, and as those citizens absorb that information, they add another layer separation to the telephone game. They might misunderstand or oversimplify what the popular writers were trying to say.

All this adds up to a great danger that even well-conducted experiments will fail to raise the level of debate on UBI. This risk doesn't require any of the people involved to be fools or fakers; this risk exists because a lot of people are involved in a long chain of transmission of very complex information. Chapter 6 argued that the 1970s discussion of the NIT experiments failed to raise the level of debate because of communication problems like these. It's important to try not to let that happen to the current round of UBI experiments.

Chapter 11: Overcoming spin, sensationalism misunderstanding, and the streetlight effect

No particular person or group is to blame for the problems discussed in Chapter 10. They result from communication barriers in long chains of interaction between very different groups trying to convey complex ideas. No complete solution exists largely because of the inherent complexity of the material and partly because of the widely varying background knowledge of the people involved.

Oversimplification is always easier to understand than real complexity.

But there is a good way to reduce this problem substantially. Everyone involved can try to understand each other better. This chapter discusses what citizens, elected officials, researchers, professional writers, and so on can do to bridge the communication gaps at each stage of the process.

The people who commission the experiment (usually elected officials) and the managers and researchers they put in charge of running it can help by consciously trying to *understand and respect the public discussion of UBI*. I have suggested that whatever else experiments do, they take on the goal of raising the level of public discussion by providing information that's useful to that discussion in a way people can best understand. I can imagine two kinds of experiments in relation to this question: one is a broad-based attempt to learn as much as possible about UBI. The other is a more tightly focused study with a more technocratic approach to certain specific questions about UBI.

Both types of studies are legitimate and useful, but neither type lies in a technocratic bubble apart from the public discussion. Other social science experiments might be free to ignore the public discussion but, as the reaction the 1970s NIT experiments (see Chapter 6) shows, UBI experiments have no such luxury. A broad-based study should make raising the level of public discussion its main goal. A more tightly focused study can still make it *one* of its goals. No matter how technocratic a particular research question might be, its relationship to the bottom line can and should be made clear, because otherwise, nonspecialists will try to make that connection themselves, making predictable and preventable mistakes (see below).

The suggestion to respect the public debate does not mean that experiments must attempt to answer every UBI-related question people might have no matter how unanswerable it might be.

But I do mean that the public discussion can be taken into account not only in the reporting of findings but also in the design of the study.

People designing a study can begin by working backward from a thorough understanding of the political economy of UBI discussion and questions important to it to the questions they can address. And then they can work forward again, explaining not just the meaning of research findings but focusing on an accurate explanation of the implications of what people most want to know.

Part three includes a discussion of claims that are important to the UBI discussion around the world, but it is no substitute for careful examination of the discussion in any particular country or region of the experiment because each local discussion is unique in important ways. If possible, people commissioning the study could even convene a forum in which supporters, opponents, and other interested parties across a broad spectrum get together with researchers to determine what claims are most at issue and what research questions would be most informative to the public discussion and public evaluation of the policy. Citizens involved in the public discussion of UBI can help by trying to get a realistic picture of what they hope to learn from an experiment rather than by making a blanket demand for a trial. Citizens' ability to do this is extremely limited because the public discussion involves millions of people who have very different political views and are not organized into a body. But writers within the movement can write about what specifically they want to learn from a UBI trial. Organizers can organize online or in-person public discussions of what people want to learn from UBI trials.

The dangers these efforts can help avoid are that people involved in the public debate will assume that experiments automatically focus on the questions of most interest to them and that specialists designing the experiment will assume that increasing the body of available knowledge on UBI automatically raises the level of public discussion on UBI. For experiments to enlighten the debate, specialists and nonspecialists need to make greater effort to understand each other.

Although the discussion varies greatly around the world, and the local political context is most important to almost any UBI study, I suggest researchers be extremely cautious about using their knowledge of the local discussion as an excuse to ignore this book's advice. Not all the claims discussed in Part Three are relevant in everywhere and additional claims will be important in particular contexts. But researchers might want to err on the side of caution by being more reluctant to subtract than to add to that list. And more importantly, I think three issues in specialist-nonspecialist understanding of UBI experiments are likely to be important in almost any political context.

First, the public discussion often conflates ethical and empirical issues. As argued in earlier chapters, researchers usually try to focus on purely empirical issues and sometimes sweep ethical questions under the rug in the process. Researchers can best separate these issues by bringing them into the open. People with different ethical perspectives are interested in different empirical claims; they are interested in different parameters to evaluate them. Framing the issue in one way or another can advantage one side or another's spin on the results, as happened with the focus on employment-levels in the 1970s experiments in the United States. A study could strive for a truly neutral framing, or it could be designed to provide information that is useful to people with different ethical perspectives relevant in the political context.

Second, people involved in the public discussion are almost exclusively interested in the long-term impact of a national UBI on almost any variable that an experiment might study. They have little or no direct interest in the simple comparison between the control and experimental groups except as an indicator of that response. Other parts of this book discuss why the difference

is often substantial in the case of UBI. Specialists conducting research for other specialists might be free to focus on simple differences between the control and experiment group. Specialists doing that for UBI experiments will not answer the questions people want answered, and they will end up misinforming any nonspecialists who attempt to interpret the results without understanding the difference.

Bridging this gap requires bringing in evidence from other sources to make predictions about how long-term community effects are likely to play out. It requires more qualitative discussion of the study's findings. It requires researchers to be unafraid of calling attention to the uncertainty of the study's predictions and to the smallness of the contribution experiments make to our overall understanding of UBI. But it is necessary to help the public discussion benefit from the contribution that experiments can make.

Third, but perhaps most important, as this book stresses throughout, people want an overall evaluation of UBI. They want a bottom line, answers to the big questions: does it work; should we do it? Although it might be naïve to believe that a UBI experiment can provide more than a small piece of the answers to those big questions, ultimately those are right questions. That's what people considering introducing a policy need to decide. No matter how tightly technically focused a research question might be, it is important to the extent that it is part of the answer to the big overall evaluation question.

Experiments cannot provide the bottom line, but they can research as many aspects of it as their techniques allow and relate all their research to it. Whatever the technical focus of the study, explaining the relationship between any particular finding and the bottom line is essential. The inherent complexity and political nature of UBI experiments makes this effort essential even though social science research conducted exclusively for others social scientists might be free from these concerns.

Respecting the importance of people's desire for a bottom line helps avoid the streetlight effect. Whoever designs a UBI experiment, might want to ask themselves: does it focus on these questions because they are the most important aspects of the overall evaluation of UBI or because they are the easiest questions to answer with the technique available? Attention to the overall public evaluation of UBI might change the focus of the study toward variables that experiments can produce only partial or indirect evidence about and toward more qualitative methods and/or methods other than RCTs.

Researchers should not neglect answering the questions trials are best able to answer, and they might have an extremely good reason for narrowly focusing their study onto issues that differ considerably from those at most issue in the public discussion, they want to avoid misunderstanding. If so, they need careful, clear answers to the questions of why they are studying what they are studying rather than what is of most interest in the public discussion and the extent to which their findings do and *do not* inform the issues in that debate. Researchers might be tempted to ignore these issues because they are obvious to other specialists in their field, but they are can be extremely difficult for nonspecialists, and these issues have historically been the source of misunderstanding.

The bottom line is also important because it forces comparison of costs and benefits and because it forces attention to things experiments cannot measure. Discussion of benefits in isolation biases the reaction one way; discussion of costs in isolation biases it the other way; even if the existence of that effect was highly predictable and the experimental question about it was merely how large it would be.

This problem was exactly the case in the 1970s experiments. Researchers focused on the labor-time effect, rightly expecting it would be negative and attempting to find out how large it was. Not everyone agrees that decreased labor hours by low-income people is a bad thing to begin with, and the experiment found effects most researchers took to be small and consistent with the affordability and certainly with the sustainability of the program. And it was an effect that was likely to be partially reversed by community effects as employers bid up wages to get workers back, creating additional positive effects of higher wages for the working poor with accompanying decreases inequality and poverty. All of these effects are things researchers can easily predict, but none of them are things an experiment—especially an RCT—can measure. None of these complexities and likely positive consequences had much of an impact on the public discussion. Many popular-media writers seized on the mere existence of any decline in labor hours as if it were not only a cost but a reason to reject the program without further discussion.⁴³

Such widespread distortion cannot reasonably be blamed on dishonesty. Most people probably sincerely misunderstood, made the predictable mistake of ignoring issues like community effects, and took one result as saying far more about their bottom line than any experimental result could say. The current round of experiments needs to head off any possible reaction like this one whether it is to the benefit of supporters or opponents. No solution is foolproof, but explicitly addressing what experimental results do and do not say about the bottom line can help.

Once the study is completed, researchers, science communication specialists, and anyone else trying to explain the experiments face all these issues in reverse. People communicating the results of UBI experiments have a more difficult job than is probably typical in other scientific communication. It is not enough simply to help people understand the experiments on their own terms—e.g. what an experiment is, what control and experimental groups are, and what differences were found between the control and experimental groups. They have to explain the relevance of those findings to the most important issues in the public discussion in ways that head off spin and misunderstanding.

And so, the communication of experimental results is inherently difficult. People doing it need a thorough understanding of the experiment, of what is important to the public understanding of this policy in the local political context, and of the many ways people are likely to misunderstand the information provided. This effort is made somewhat easier because many common errors are predictable. For example, whether because of sensationalism or professional deference, some people are likely to see experimental results as more inclusive than they are. Whether because of a desire to spin or overconfidence in the meaning of research, some people are likely to discuss various results out of context as if they were votes in favor or against the adoption of UBI nationally.⁴⁴

These kinds of misunderstandings and misuses can be partly headed off by clearly discussing what the experimental findings do and do not imply about wider issues. For many issues, this effort involves admitting that the experiment provides indirect evidence that has to be combined with other evidence to make the necessary predictions. For some issues, it requires admitting the experiment produces no evidence.

People directly involved in the experiments are not the only ones who can help create a better public understanding of UBI experiments. Anyone with good knowledge can help raise the level of public understanding making themselves heard—and understood—to counteract any spin and misreporting.

Outside researchers who understand the place of experiments in the political economy of the UBI debate can reexamine and represent findings in ways they recognize as more useful to the public discussion and less likely to be vulnerable to spin or sensationalism.

Journalists, bloggers, and anyone interested in writing about UBI trials can help before and after the experiment if they take time to investigate these difficult issues and avoid the easy but sensational simplification.

Citizens—it could perhaps go without saying—can help by exploring the diverse literature that will be produced on UBI experiments and reading it critically.

Part Three: From the debate to the test

Chapter 12: The bottom line

Part Two above argues that UBI experiments are more likely to succeed in the goal of raising the level of public discussion if researchers work backwards from an understanding of the place of UBI experiments in the public discussion of UBI (identifying the claims that are important to the public discussion and relating all their findings to those claims) and forward again, explaining the relevance of the experimental findings to the issues that are important to the public discussion. This effort does not require experiments to test everything everyone wants to know about UBI or even most of the claims discussed in Chapter 13, but it does mean that the experiments will be better understood if researchers relate all of their findings to the issues that are most valuable to the public evaluation of UBI as a policy, admitting the limits of what experiments can say about these issues.

Part Three now takes on the effort of working backward from public discussion. This chapter attempts the first step in that process, identifying a bottom line. The following chapter attempts the next step, identifying more-specific claims that are important to the discussion. The following three chapters discuss the extent to which experiments can directly or indirectly address each of those claims. After that, the book considers how to frame testable research questions that are relevant to those claims and considers options for designing experiments in light of them.

Previous chapters have also argued that public discussion of almost any potential public policy has a strong interest in an overall evaluation in both empirical and ethical terms. Does it work? Should we do it? Experiments are empirical studies that cannot address purely ethical claims at all. Experiments will have to focus on some aspects of the does-it-work question. Unfortunately, separating that question from the should-we-do-it question is more difficult than it might appear, because controversial ethical judgments have to be made before we can meaningfully address some of the most important empirical questions, as the following paragraphs explain.

The question, *does UBI work*, is too vague for a social science experiment, partly because whether something “works” depends on controversial ethical questions such as what goals it is supposed to accomplish and should we weigh various goals relative to each other and to various negative side effects. Whether something (such as a decline in average labor hours among low wage workers) is considered a negative “side effect” or a positive “effect” often depends on controversial ethical issues.

Social scientists tend to translate the does-it-work question into the question: how cost-effective is it? This question sounds very scientific and neutral, but it still requires a resolution to controversial ethical questions. Which effects of UBI morally count as costs? Which count as benefits? What relative weights do we put on benefits X, Y, and Z in a comparison to each other and to costs A, B, and C, and what weights do we put on them? If citizens and policymakers could resolve all of these issues and hand empirical researchers an index to weigh costs and benefits, researchers would have a purely empirical question to examine. But no one can resolve these deep moral controversies in advance of a study.

Empirical researchers are, therefore, forced to impose some controversial judgments on their evaluation process. They should at least warn readers what these judgments are. That is, they can add yet another caveat. Perhaps, they should go farther and examine several different moral weighting systems to provide information for people with diverse ethical positions.

Empirical economists sometimes ignore the ethical background assumptions in their evaluative tools. Many economists look at costs exclusively in dollar terms and cast cost-benefit calculations in efficiency terms with little or no discussion of the debate over whether these measures should have ethical priority over other options. For example, although a dollar lost to anyone is an efficiency loss, citizens might have good ethical reasons to value a dollar used to cure poverty more than several dollars used to provide luxuries for the already wealthy.⁴⁵

In the absence of a national resolution to the ethical controversies that create this problem, researchers will have to impose something. However, they should avoid presenting their resolution to moral issues as if it were uncontroversial because that would be highly misleading, and it would bias the political reaction toward that moral point of view. It is better to be open about the kinds of moral judgments that need to be made to frame the empirical issues. It's also valuable to recognize the different moral perspectives that are relevant in the local political context, and present evaluations relevant to each. This book cannot resolve this issue and won't dwell on it. I bring it up just to make sure people are aware of it before I attempt to identify a reasonably neutral bottom line.

The best research-relevant approximation of the does-it-work question might be the cost-effectiveness question. I attempt to state a UBI cost-effectiveness question broadly as:

Is a fully implemented national UBI a cost-effective method to benefit people in the short and long run in the ways UBI supporters claim it does, assuming cost-effectiveness is judged relative to other methods of achieving similar benefits for the same people?

This question could alternatively be stated: What policy (basic income, the current system, or any other alternatives to be tested) produces the greatest increase in recipients' welfare per unit of cost (both in terms of tax cost and efficiency loss), in the context of a long-term, fully implemented national policy?⁴⁶

Obviously, these statements of the bottom line can be shortened if some of their constraining phrases can go without saying. I hesitate to do so because of the amount of misunderstanding these issues have caused in the past.

I suggest that the cost-benefit question—or something like it—should be considered the bottom line for UBI experiments. Experimental evidence cannot provide the bottom line, but this chapter considers how researchers can relate experimental findings to it.

My specifications of the bottom line impose answers to some moral questions. I've tried to reduce this problem by phrasing the question in relative terms—relative to supporters' claims

about its benefits and relative to other ways of achieving those benefits. It intentionally leaves open what the claimed costs and benefits are. The next chapter discusses some popular claims, and these will vary by nation or region. I'm concerned with over-identifying any claim as "the" goal of UBI in any political context. The UBI movement is diverse. Some see it as a way to eliminate the threat of poverty for everyone. Some see it as a way to make alternative lifestyles possible. Some see it as a way to simplify and streamline the tax and benefit system. And so on. I doubt there is any political context in which virtually everyone who supports UBI supports it for the same reasons.

Phrasing the cost-effectiveness question in relative terms does not eliminate moral controversies. For example, even if nearly anyone might agree that a central goal of UBI is to "increase recipients' welfare" (as used above), any effort to define "welfare" is controversial. Popular welfare measures might leave out some of the concerns that are important to the UBI movement. Researchers should not, therefore, stop using them, but they can supplement their use of them by addressing how UBI affects items that are left out but important to the public discussion.

I'm sure there are many ways to improve my phrasing of the cost-effectiveness question, and what is best will vary by political context. The more important point is not that the bottom line is phrased this way. More important points are that the experiments have a bottom line, that it is a broad question, that it compares costs and benefits, that it focuses attention on things experiments cannot measure, and that it addresses what people need to know to evaluate UBI as a potential national policy.

I mean two things by identifying a question as the bottom line. First, virtually any empirical research question can (and probably should) be understood as some part of the answer to this general question. Second, this question, or something like it, is what citizens and policymakers ultimately want empirical research to tell them about UBI's effects, if they are going to evaluate it as a policy. The more they know about the cost-effectiveness of UBI, the more fully informed they will be as they discuss and make the decision whether to implement UBI. If citizens and policymakers believe many of the media reports on the launch of experiments, they are *expecting* an answer to this question or even to a bigger question incorporating ethical issues as well.

An experiment has much a narrower objective than what people might expect. Experiments divide people into control and experimental groups, observe whatever differences they can, and test those differences for statistical significance. If experimental reports are limited to explaining what these differences are, they stop very short of the bottom line that people are interested in. This is fine when specialists communicate with other specialists who understand the limits and implications of those limited findings, but it is not of much value to the public discussion and causes predictable problems discussed in chapter 11.

No one evaluating UBI as a policy option has a direct interest in the difference between control and experimental groups for any variable. They have an interest in a long-term estimate for the impact of a national UBI on that variable. And they have an interest in viewing it in the context of cost-benefit analysis relative to other policies. Consider that the bottom line for any particular variable. Calculating the bottom line for any particular variable involves considering the long-term impact of a national UBI on that variable. This calculation involves considering community effects, the difference between a short-term study and a permanent policy, the ways in which the sample is unrepresentative of the entire population, and so on. For some variables, researchers might be able to use simulation techniques to calculate that answer. For others, they might have to bring in more qualitative data or simply have a qualitative discussion. Even if they lack data to make a reasonable estimate, they can explain the differences between what they found

and what we really want to know, and they can discuss the missing factors they would need to be estimated to get closer to the bottom line.

Calculation of any overall bottom line requires a comparison of the bottom line for each particular variable estimated in the experiment and probably also with estimates for other variables the experiment could not examine. This effort, again, might be achieved with simulation techniques, might instead require more qualitative techniques, or might involve admitting why the effort falls short of that goal.

All of this effort calls attention to what experiments cannot do. But it is necessary to help people truly benefit from what experiments can do.

Chapter 13: Identifying important empirical claims in the UBI debate

This chapter presents two lists of claims that supporters and opponents have made about the effects of UBI. It gives each claim a name for easy reference, but these names do not reflect any standard definition. I have tried roughly to group similar claims together, going from the more common or important groups of claims to the less common or less important groups of claims, but the order is not terribly important. My estimates of how best to group claims and of frequency and importance of claims is cursory and subjective. And of course, the importance of any claim varies place-to-place and over time.

The list includes a definition for each claim, but not any further discussion, such as how it is supposed to work. Later chapters give further explanations as needed.

There is some overlap between the claims. I have tried to include overlapping claims only when related claims play important, individual roles in the debate. Many claims could be divided into a series of more-specific claims. The welfare claim and the cost-effectiveness claim are obvious examples. Only some of the more-specific claims are included separately on the list.

It would be possible to match almost any claim on one list with its negation on the opposite list. For example, supporters tend to say UBI is cost-effective and affordable while opponents tend to say it is unaffordable and cost-ineffective. I have tried to avoid these sorts of duplications by attributing it to the side that seems to focus on the issue more. Therefore, the supporters' list gets a cost-effectiveness claim but the opponents list gets an (un)affordability claims.

Although the lists don't claim direct negations, they do include some pairs of opposite claims. For example, the benefit-to-workers claims and the harm-to-workers claims are included on the two lists because they play important separate roles in common arguments for and against. Many supporters don't stop at defending against the allegation that UBI harms workers; the ways in which UBI is likely to help many workers play an important role in arguments for UBI.

Not all supporters or all opponents agree with each of the claims on the respective lists. In fact, some of the claims within each list contradict each other. This is to be expected, given that diverse people support or oppose UBI for many different reasons, and sometimes have little in common but their support or opposition to a certain type of policy proposal.

These lists are not meant to exhaust all reasons given for against UBI. Based on my experience, however, they capture a large portion of the common and influential claims in UBI literature. I expect that all or most of the questions experiments examine are related to some of the claims on these lists.

Claims commonly made by supporters

- The welfare claim: UBI significantly raises the welfare of net recipients and many net contributors.
- The poverty claim: UBI (usually in combination with other policies) can eliminate poverty.
- The benefit-to-workers claim: UBI financially benefits many workers directly by acting as a wage subsidy for lower-income workers and indirectly by creating market conditions likely to increase wages.
- The better-working-conditions claim: UBI improves working conditions for many workers both by giving them the flexibility to move to more attractive sectors and by creating market conditions likely to give employers incentive to improve working conditions.
- The widespread-benefit claim: a large portion of the population will benefit (on average) from UBI at any one time, and a substantially larger portion will benefit at some point in their lives.
- The flexible-lifestyle claim: UBI enables people to work shorter hours, engage in job sharing, become fulltime parents, and so on.
- The freedom claim: UBI gives people greater freedom in the sense of giving them more effective power over their own lives by reducing or eliminating their dependence on employers.
- The anti-exploitation claim: UBI reduces exploitation in employment by giving all workers the power to refuse exploitive working conditions.
- The economic-equality claim: UBI increases economic equality both by direct redistribution to lower-income people and by indirect effects, such as creating more favorable labor-market conditions, improving health, and increasing education. (The taxes used to support it can also be formulated to increase equality.)
- The social-equality (or social-inclusion) claim: UBI increases social equality by reducing social isolation of low-income people, by reducing the stigmatization of people who benefit from redistributive programs, by reducing housing segregation, and by other means.
- The economic-and-social-mobility claim: UBI increases economic and social mobility by improving the health, security, and education of children, and by helping adults start businesses, get education or training, take the time to look for the right job, and so on.
- The cost-effectiveness claim: UBI is relatively more cost-effective than traditional, conditional welfare policies (in achieving various goals).
- The reduced-social-costs claim: by reducing poverty and inequality, UBI reduces costs associated with them such as healthcare costs, policing costs, and so on.
- The reduced-capture-corruption-and-bureaucracy claim(s): UBI reduces the overhead cost associated with income support. Its benefits are less likely to be captured by others (such as employers, landlords, and bureaucrats) than conditional welfare state policies. And it is less vulnerable to corruption than conditional programs (because of its simplicity and transparency).
- The efficient-transfer claim: UBI, being a lump-sum transfer, is economically efficient. The only social cost involved with it comes from increases in marginal tax rates associated with financing it not from the grant itself.

- The poverty-trap claim: UBI encourages people on benefits to reenter the labor force in greater numbers than a conditional system, by ensuring they are always better off earning more private income than earning less.
- The labor-productivity claim: UBI increases labor productivity by encouraging employers to substitute skilled for unskilled workers, by improving workers' ability to enhance their skills and search for higher-productivity jobs, and by improving childhood health and educational attainment.
- The increased-innovation-and-entrepreneurship claim: UBI increases entrepreneurial activity and innovation (because it gives people a financial cushion to take risks, and more time and more investment capital to pursue their ideas)
- The productive-non-labor claim: UBI allows people to do more unpaid work (such as care work and volunteering), some of which is more productive (or socially valuable) than many forms of paid labor
- The increased-support-for-redistribution claim: Once in place, UBI results in greater overall political support for redistribution.
- The politically-enabled-proletarian claim: UBI—makes low-wage workers a greater force for progressive social change on other issues by freeing them from long hours and low pay.
- The economic-stimulus claim: UBI, in combination with the taxes that support it, helps improve economic growth and reduce unemployment by helping to stimulate and stabilize aggregate demand.
- The “degrowth” claim: UBI helps economies move away from overconsumption and overexploitation of resources.

Claims commonly made by opponents

- The reciprocity (or work ethic) claim: UBI makes it possible for non-wealthy people to share in the benefits of social production, which involves labor, without making a reciprocal labor contribution of their own—or without any meaningful social contribution at all. This observation is often labeled a violation of norms such as reciprocity and/or the work ethic.
- The exploitation claim: UBI requires taxing the income of workers for the benefit of nonworkers.
- The harm-to-workers claim: a UBI system financially benefits nonworkers at the expense of many workers, all effects considered.
- The labor-effort claim: UBI causes an unacceptably large reduction in labor supply that is not easily counteracted by other policies.
- The affordability claim: UBI at the proposed level is prohibitively expensive.
- The economic-impediment claim: UBI decreases economic growth by various means including, reducing labor market participation, increasing labor costs, causing inflation, creating the need for increased taxes, which reduce investment and innovation, and so on.
- The self-destruction claim: UBI increases self-destructive behavior (possibly including drug-dependency, lack of care for the future, watching too much television, playing too many video games, choosing meaningless activities over meaningful paid work, having “too many” children, etc.).

- The gender-role reinforcement claim: UBI helps maintain traditional gender roles by making it easier for women to remain out of the paid labor force while performing unpaid care work and other traditional women’s roles.
- The consumerism claim: UBI, being a cash grant in a monetary economy, encourages greater consumerism, leading to increased environmental destruction and other problems.
- The bought-off-proletarian claim: UBI—by providing a minimal level of contentment for workers—reduces their effectiveness as a force to challenge the deeper inequalities and other social inequities in society.
- The decreased-overall-redistribution claim: UBI at an economically or politically feasible level makes low-income people worse off overall than traditional, conditional social policies.
- The capture claim: many of the benefits of UBI go to someone other than the recipients (perhaps because employers reduce wages, because the cost of housing in low-income areas increase because bureaucrats create overhead costs, etc.).
- The migration claim: UBI encourages immigration and/or migration into areas with UBI.
- The shut-door claim: UBI creates political pressure to restrict immigration.

Conclusion

These lists are bound to be incomplete. I’ve compiled them from general but far-from-perfect knowledge of the UBI debate accumulated over years. Many more claims (of various levels of relevance, certainty, and testability) are undoubtedly circulating in the academic and nonacademic literature on UBI. But I hope these claims captures a significant range of what is being said. The diversity of claims on these lists is enough to demonstrate the difficulty of designing and communicating the results of a UBI experiment in a way that successfully raises the level of discussion.

The next step is to consider how much an experiment can say about these claims and what research questions are useful in light of the concern about these claims.

Chapter 14: Claims that don’t need a test

At least four of the claims on the lists above don’t need a test to confirm their truth as stated. They are true either by definition, or they can be shown to be true by analytical reasoning with little or no empirical reasoning necessary. These include,

- The poverty claim
- The freedom claim
- The efficient-transfer claim
- The reciprocity (or work ethic) claim

These claims are related to important claims that can be researchers, and they can be used to help frame related research questions, but it is important to understand that they play a prominent role in the UBI discussion as stated.

The efficient-transfer claim is analytically true. All lump-sum transfers are efficient in the sense defined by economists. That is, recipients benefit by the full value of the grant and not

by a lesser amount. Non-lump sum transfers give individuals incentives to change their behavior to get the grant. These changes in behavior cause additional social costs. This fact has played a prominent role in the discussion among economists since James Buchanan, F. A. Hayek, and James Tobin endorsed the idea. Lump-sum taxes, such as those on resource rents, are also efficient, and if UBI can be financed entirely by such taxes, no social cost would be involved with UBI at all. Experiments cannot test whether lump-sum taxes can raise enough revenue to support UBI, and experiments will probably assume that a substantial increase in marginal tax rates will be necessary to finance UBI. All the costs attributable to the size of UBI will be an efficient transfer. Only those costs associated with the marginal tax rates affecting recipients constitute an efficiency loss.

Two research questions closely related to efficiency are extremely important: *what portion of UBI's cost represents an efficient transfer and what portion represents a true social cost? How does the efficiency loss of UBI in these terms compare to the efficiency loss of an equally generous expansion of existing programs?* These two questions have been neglected by most past experiments. The labor market findings of UBI experiments will be useful toward answering these questions, but it will have to be the experimental findings will have to be combined with a large amount of outside evidence to produce a result.⁴⁷ The need for evidence from other sources will be a running theme as these chapters try to relate the questions people want answered and the questions experiments can directly examine.

The poverty claim as stated (the statement that UBI can eliminate poverty) needs no test. Poverty is defined as having less than the poverty level—an amount of income judged as enough to afford basic necessities. A UBI set at or above that level necessarily eliminates poverty. UBI's ability to *eliminate* poverty is an important advantage over the traditional approach because virtually all conditional systems leave some portion of the population in poverty. If the conditions are to be credible threats to get people to do things they might not otherwise be willing to do, conditional programs would *have* to leave some people in poverty.

Strictly speaking, the poverty claims not true by definition because it is possible that a poverty-level UBI is economically unsustainable. Existing evidence overwhelmingly indicates that a poverty-level UBI is sustainable in high-income countries (see the affordability claim discussed below). Of course, it won't hurt to double-check the sustainability of UBI, but the sustainability of a poverty-level UBI is not a pressing source of serious disagreement in the debate.

A sensational media headline saying *study finds poverty elimination possible with UBI* would be true, but it would not meaningfully report a groundbreaking finding. Such a headline would spin the discussion of research findings to the pro-UBI side. Yet, leaving UBI's ability to eliminate poverty out of the discussion of the findings spins the issue to the anti-UBI side.

The poverty claim is useful in framing research questions around the cost-effectiveness claim. The question, *what is the cost of eliminating poverty with a UBI*, is fairly neutral. But a non-comparative focus on cost creates a spin opportunity for the anti-UBI side. Perhaps a more neutral way to approach this issue is with a two-part questions: *What are the relative cost and effectiveness of attempting to eliminate poverty with a UBI or with increases in existing transfer programs?* There are other ways to look at relative cost-effectiveness, but certainly, its relation to UBI's ability to eliminate poverty is one important way. See the discussion of cost-effectiveness below.

The freedom claim and the reciprocity claim (as stated) are true by definition. The controversy is not over their truth but over their moral content. UBI set at a sufficient level, undoubtedly gives non-wealthy people greater control over some aspects of their lives, increasing freedom in the sense used in the freedom claim. The same UBI makes it possible for non-wealthy people to consume products that involve labor without themselves contributing labor, violating the

reciprocity principle in the sense used in that claim. No empirical investigation can settle the disagreement over the moral value of these senses of freedom and reciprocity.

There are important closely related empirical questions. The extent to which the benefit-to-workers claim, the productive-non-labor claim, and the flexible lifestyle claim hold true would indicate something about how valuable the added freedom for low-income people was, but unfortunately, UBI experiments are not the type of empirical study necessary to investigate them (see below).

UBI experiments can contribute something to the question of whether more people violate this reciprocity principle under UBI, capitalism as is, or under an expanded conditional welfare system. However, to do so they would have to define the ethically controversial concept of meaningful social contribution. It would be problematic for them to make such an attempt and many people would object to whatever definition they chose.

One of the most valuable things researchers can do about the reciprocity issue is to head off the interpretation that experiments say more about it than they do. Experiments can and will certainly collect data on the labor time of the control and experimental groups. Opponents are likely to interpret any decline in labor time as an indication of a violation of the work ethic, and some writers are likely to spin it as such, as many did with the results of the 1970s NIT experiments.⁴⁸ Merely presenting labor-time findings—even on the way to calculating its effect on cost—without addressing its possible effect on the reciprocity principle invites that misconception among people for whom that principle is a primary concern. To head off that mistake, researchers will have to pay close attention to that concern and assess whether any labor-time decline reflects people dropping out of the labor force or merely reducing the number of hours the work. If researchers stop there, they leave open the interpretation that work is the only meaningful social contribution. But if they go farther, they step into the trap of trying to defining controversial moral claims. They can discuss the issue conceptually without getting into specific estimates of what should count, but some confusion on this issue might be inevitable.

Furthermore, even if experiments could somehow show that UBI was very unlikely to cause any increase in violations of the politically relevant versions of the reciprocity principle, the truth that UBI makes it *possible* for non-wealthy people to live without laboring is likely still to feature prominently in the debate. People who are strongly motivated by the reciprocity principle might prefer to eliminate the possibility that anyone violates it. Their moral judgement might be that non-wealthy people are not working enough now and that only a program that increases how much they work is acceptable. People who hold such beliefs are unlikely to find any experimental data useful.

Some spin and some misunderstanding on all of these issues is inevitable. The goal is simply to reduce it as much as possible. To do so, anyone writing about experimental results needs to present it in a way that answers people's questions about how the findings relate to these issues. Few if any nonspecialists will be able to work out many of these issues for themselves, and they won't be helped much by a dry list of caveats.

Chapter 15: Claims that can't be tested with available techniques

Many important, empirical claims in the UBI discussion are untestable or virtually untestable by the techniques available to potential UBI experiments. These include the following.

- The exploitation claim

- The anti-exploitation claim
- The social-equality claim
- The capture claim
- The reduced-capture-corruption-and-bureaucracy claim(s)
- The labor-productivity claim
- The increased-innovation-and-entrepreneurship claim
- The economic-and-social-mobility claim
- The better-working-conditions claim
- The flexible-lifestyle claim
- The productive non-labor claim
- The gender-role reinforcement claim
- The degrowth claim
- The consumerism claim
- The self-destruction claim
- The economic-stimulus claim
- The economic-impediment claim
- The migration claim
- The shut-door claim
- The increased-support-for-redistribution claim
- The increased overall disadvantage claim
- The politically enabled proletarian claim
- The bought-off proletarian claim

The anti-exploitation claim and **the exploitation claim** are not polar opposites. The anti-exploitation claim involves exploitation of workers by employers and UBI's suspected ability to reduce it. The exploitation claim involves exploitation of workers by non-working UBI recipients. Depending on how exploitation is defined, it is possible for both claims to be true at the same time. A UBI could make workers less vulnerable to exploitation by employers while making them more vulnerable to exploitation for the benefit of non-working net recipients. Similarly, the exploitation claim is distinct from the harm-to-workers claim. The exploitation claim focuses only on the effect of taxes. It is possible that some workers pay higher taxes under a UBI system, and so are exploited, but are better off overall because of better wages and working conditions, as well as other community effects. See the discussion of the harm-to-workers claim in the next chapter.

The concept of exploitation is so controversial and so morally loaded that researchers can't hope to say much about it directly, but it is so important that they should not ignore it either. They need to address other issues, such as the welfare claim, the benefit-to-workers claim, the better-working-conditions claim, and the harm-to-workers claim in the context of the exploitation debates. Unfortunately, these are difficult to address as well, as discussed in the following chapter.

Although **the social-equality claim** is often given as an important reason for UBI and other forms of redistribution, experiments can say very little about it, because it is inherently a community effect. Experiments will not directly reveal whether UBI net beneficiaries are less likely to be stigmatized than recipients of other redistributive programs. They won't observe housing segregation. Experimenters can ask people whether they feel socially isolated, but any relief from isolation is likely to be much larger in a long run community-wide experiment.

The capture claim and the reduced-capture-corruption-and-bureaucracy claim(s) cannot be tested in an experiment because they involve market reactions and/or the internal workings of a potential future government administration. Small-scale, temporary trials are administered in very different ways than large-scale national bureaucracies. Therefore, the bureaucratic structure needed to run an experiment will provide no evidence about the bureaucratic structure needed for a national program or about the behavior of public employees within that structure. To the extent that these claims involve capture by private economic entities such as employers and landlords, an RCT will provide no direct evidence and a saturation study will provide very little. Labor markets are national. The effect of geographically dispersed randomly selected individuals will be nonexistent. The effect of geographically concentrated subjects in a saturation study will probably be much smaller than the national response, and how large it is will depend on how isolated the community is.

The labor-productivity claim, the increased-innovation-and-entrepreneurship claim, the economic-and-social-mobility claim, and the better-working-conditions claim are extremely hard to observe because they depend on the long-term reactions of both recipients and employers. Researchers can examine whether people in a short-term experiment seek training or education, whether they are healthier, and so on, but they will be unable to observe whether and how any gains in these areas will eventually affect workers' productivity, entrepreneurship, and mobility. A major part of the argument for increased labor productivity and improved working conditions is through employers: a decline in labor effort gives employers incentive to increase wages, improve working conditions, and introduce higher productivity techniques. Because RCTs are unable to observe any employer responses, they cannot observe whether this path actually leads to higher productivity or better working conditions. The best they can do to approach employer reaction is to observe whether the UBI trial leads to a decline in labor market participation, which is only the first step in the chain expected to lead to these results.

The flexible-lifestyle claim, the productive non-labor claim, the gender-role reinforcement claim, the degrowth claim, the consumerism claim, and the self-destruction claim all share two problems. They require observing behavior that is not easy to observe and making subjective and/or normative judgments about that behavior. For example, researchers can observe whether parents use their UBI to spend more time with children and whether women do this more often than men, but they will not be able to observe whether this reaction should be seen as reflecting increased flexibility in lifestyles or as reinforcement of a gender role. It will be very difficult to observe whether test subjects react in ways that lead to more or less growth and consumerism, and even if researchers are able to observe what subjects do with increased available leisure time (should they choose to take it). Researchers would have to make controversial moral judgements to label that time "productive," "unproductive," "self-destructive."

Yet, researchers will need to find some nonjudgmental way to make findings about subjects' behavior relevant for these debates. For example, although they should avoid making moral judgments, they should not avoid attempting to see whether UBI is correlated with alcohol or drug abuse.

In addition, most of these variables are likely to depend heavily on long-term and community effects. For example, the ability of a person using a UBI to adopt a more flexible lifestyle is likely to depend on factors such as whether the UBI is permanent, whether it affects the market and culture in ways that make flexible lifestyles more feasible and attractive. Any short-run observations of people in a small-scale experiment are likely to give little indication of the long-run reaction to a national UBI for any of these possible effects.

The economic-stimulus claim, the economic-impediment claim, and associate sub-claims are all claims about market reaction to UBI, which RCTs cannot observe at all and saturation studies can only observe partly. Some of the potential effects involved are macroeconomic, operating only at the national and—in the case of the Eurozone—at the supranational level. A small-scale experiment can say nothing about them. Evidence has to be gathered from other sources.

The migration claim, the shut-door claim, the increased-support-for-redistribution claim, and the decreased-overall-redistribution claim are claims about how voters and policymakers feel about UBI and respond to it at the national level over time. Despite their potential importance to the debate, small-scale experiments cannot provide any evidence about them.

The politically-enabled-proletarian claim and the bought-off-proletarian claim are potentially observable in an experiment because they are claims about the behavior of people in response to receiving a UBI. Researchers could examine the political behavior of test subjects, and try to identify whether they are more or less politically active than the control group. But that there are at least four reasons researchers are probably better off admitting this question is beyond the reasonable capability of an experiment. First, political behavior is extremely difficult to observe and hard to quantify. Second, community effects are likely to be substantial. The way one person behaves politically affects how their neighbors and fellow citizens behave. A RCT would capture none of that effect, and a saturation study would capture only a little of an effect that might be much larger for a national program. Third, once a national UBI is in place, it would change the political dialogue in unpredictable ways, and the behavior of people would respond in unpredictable ways. Their behavior in an experiment will give no indication of that response. Fourth, the long-term political response after years of activity and discussion in a particular policy setting is likely to be very different from the initial reaction of study subjects.

Although I do not recommend that researchers attempt to answer these questions, researchers should be aware that these claims affect how people interpret the results that experiments are able to produce. Suppose experiments show that the experimental group (receiving the UBI) work fewer hours than the control group. This result could be a good thing because it is the first step in a process consistent with the anti-exploitation claim, the better-working-conditions claim, the reduced-capture claim, the labor-productivity claim, the productive non-labor claim, the degrowth claim, the capture claim, the consumerism claim, and the politically-enabled-proletarian claim. But this result could be a bad thing because it is the first step in a process consistent with the exploitation claim, the gender-role reinforcement claim, and the economic-impediment claim. People who feel strongly about these issues are likely to see confirmation in the results, glossing over the distance between the first step that might be confirmed by the experiment and the final step required for their theory to produce the result they expect. Keeping people from making this leap is a difficult challenge for anyone writing about experimental findings.

The difficulty of relating the trial findings to the issues being debated might tempt researchers to report experimental results on their own terms without any comment on what they indicate for all these different debates, but as the 1970s experience shows, ignoring these debates makes it easier for the two sides involved in them to spin the results one way or another. It might be better for researchers to recognize these debates and explain—in neutral language—the limited implications experimental results have for these debates.

Chapter 16: Claims that can be tested but only partially, indirectly, or inconclusively

Experiments have some ability to examine the following claims, if only partially, indirectly, or inconclusively.

- The welfare claim
- The reduced-social-costs claim
- The labor-effort claim
- The affordability claim
- The poverty-trap claim
- The economic-equality claim
- The harm-to-workers claim
- The benefit-to-workers claim
- The widespread-benefit claim
- The cost-effectiveness claim

The welfare claim

The welfare claim is probably the most important empirical claim in the UBI debate. The central reason to support a transfer payment is to make people better off. Although some past studies have underplayed the welfare question in favor of more easily measurable variables, the question of how well UBI achieves that goal is primary in importance. Research on a topic like this needs to focus on effects more than side effects.

Welfare is an abstract concept about people's inner state and is not directly observable. The best existing methods for determining welfare are asking people and observing quality-of-life indicators.

Some controversy exists about quality-of-life indicators. Welfare is at least partly subjective, and some indicators can be morally loaded. Alcohol is clearly unhealthy and has many potentially damaging side effects, but if it has no ability to increase welfare, 70% of Americans—some of them very well-educated and refined—don't know what's good for them.

Fortunately, many quality-of-life indicators are not as tricky. If you have more secure access to an adequate diet, more secure housing, fewer feelings of social isolation, and healthier, longer-lived children than otherwise, you are almost undoubtedly happier. People who are made happier by an inadequate diet, shorter-lived children, and so on probably suffer from a diagnosable mental disorder. And so, we can safely use many quality-of-life indicators.

Social scientists have developed reasonable welfare indexes based on well-researched lists of quality-of-life indicators. Researchers conducting UBI experiments can report on quality-of-life indicators in a nonjudgmental way and employ various respected quality-of-life indexes to provide an overall measure of welfare. They can also conduct a survey asking people in the control and experimental groups about their wellbeing and about factors likely to affect it.

The need for welfare indicators means that the welfare claim is really a host of claims, or sub-claims, if you prefer that term. I haven't attempt to list each claim separately, because there

are just too many of them, including effects on physical and mental health, homelessness, housing quality, infant mortality, education, food security and adequacy, nutrition, problems associated with high concentrations of low-income people (the “ghettoization” of poverty), and many more.

Researchers could straightforwardly employ standard quality-of-life indicators and welfare indexes, but they might also consider addressing welfare issues that have particular importance to the UBI debate, such as those related to the freedom claim, the flexible-lifestyle claim, the consumerism claim, and the self-destruction claim. Chapters 14 and 15 discussed the difficulty of dealing with these claims at all, and they get into the controversial territory discussed at the beginning of this chapter, but these issues do affect welfare, and they have particular importance to the UBI discussion in many countries.

The sheer volume of welfare indicators distracts attention from how important each of them is. I’m guilty of that myself, leaving them out of the list of named claims. But UBI experiments must emphasize all quality-of-life indicators they can measure and explain the relevance of the ones they can’t.

The difficulty of observing, measuring, quantifying, and combining quality-of-life indicators into a welfare measure also distracts attention from the welfare question. Information about welfare indicators can be gathered from observation of recipients purchasing and behavior and from surveys of recipients about their behavior, attitude, and feelings. Doing so is highly imperfect and involves a great deal of subjective judgement. But doing it as well as possible is essential if we’re going to examine whether UBI succeeds in doing what it is proposed to do—improve the wellbeing of net recipients. By contrast, the labor-time comparison between the control and experimental groups is one nice, neat number that people can easily read and (apparently) interpret. Allowing attention to fall on the more quantifiable variables is an example of the streetlight effect.

As always, researchers are limited to studying quality-of-life indicators of the control and experimental groups over the course of the trial, when we really want to know the long-term effects of UBI on a national level. Community and long-term effects are likely to be substantial because UBI is likely to affect welfare through many different channels through direct distribution, through market effects on income and working conditions, by reducing inequality, by reducing the ghettoization of poverty, by improving education, and so on. Researchers will have to do a great deal of extrapolation to relate the two, bringing in more and more uncertainty. Individual-level RCTs, being unable to measure any market reactions or community effects of UBI will underestimate the impact of UBI on quality-of-life indicators—both positive and negative. A saturation study, with limited ability to measure market reactions and community effects, is likely to underestimate these indicators to a lesser degree. Both kinds of studies suffer from the problem that the effect of UBI on many quality-of-life indicators is likely to accumulate slowly over the long term, to be larger for a policy expected to be permanent, and to involve national-level community effects.

One advantage of saturation studies is that many welfare-related community effects are local. Evidence indicates that it is far more damaging to be poor in a neighborhood where everyone is poor than it is to be poor in an economically integrated neighborhood. A 5-to-10-year saturation study in an impoverished town—if feasible—could produce a great deal of information about the effects of ghettoized poverty not just about UBI’s role in alleviating it.

The trial will give some indication about the direction of UBI’s impact on various welfare indicators, but researchers will have to extrapolate using other evidence to estimate the welfare impact of a national UBI including the feedback effects from employers and the community, and

over the long term. And once again those predictions will be based largely on that other evidence and only partly on the experimental evidence.

However, if the short-term experimental impact is large and in one distinct direction, then it is reasonable to believe that the long-term impact of a national UBI will be larger and in the same direction. So, although the experimental results can't be conclusive, they can provide important information.

The reduced-social-costs claim

Experiments can address **the reduced-social-costs claim** by examining the demand for social services among experimental subjects. UBI's potential to alleviate the poverty trap is only one example. Not all social costs are observable this way, and so the results will be only partial. Experiments cannot reveal the full impact of UBI on the demand for social services because that demand greatly depends on community effects, and they are likely to be larger in the long run. Researchers will have to rely on non-experimental evidence to fill in all these gaps to estimate the effect of UBI on social costs. Experimental findings of the demand for social services will play only a small role in that estimate.

The poverty-trap claim

The poverty-trap claim is that UBI will lead to greater labor effort for people eligible for full-time benefits under a conditional system. This can happen because many conditional programs (such as disability, public housing, unemployment insurance, and in the United States, free or subsidized medical care) require people to sacrifice all or most of their benefits if they accept employment or have private income above a certain level. This rule gives recipients a financial incentive to choose benefits over low-paid labor, possibly preventing them from taking the first step toward economic mobility—hence the “trap.” Some conditional programs have effective marginal tax rates in excess of 100%, so that recipients are financially better-off remaining on benefits than they would be taking a low-wage job.

UBI eliminates the poverty trap because people receive the grant regardless of income. Depending on the structure of the program, higher income will probably be accompanied by a larger tax bill either directly or indirectly, but not at rates near 100%.

A UBI experiment can test reasonably well whether people—in the short term—respond to the removal of the poverty trap at a given wage. Of course, an RCT cannot test whether UBI *also* creates market conditions likely to increase wages and improve working conditions further enticing people into the market. Nor can it determine whether improvements in health, education, housing, food security, and similar variables will increase people's ability to get out of poverty in the long run. A saturation study's ability to estimate these community effects is limited, and additional theory and evidence will have to be combined with experimental findings to produce an estimate.

For this issue, it is extremely important to separate the effects of the size of transfer from the effects of the type of transfer. If a large UBI is tested against a small conditional program some or all of the work-stimulating impact of removing the poverty trap will be counteracted by the creation of a more generous alternative to work.

The labor-effort claim

Experiments can provide some direct evidence about **the labor-effort claim**, but it is deceptive. Experiments will observe the difference between the average number of hours worked by the control and experimental groups, and that comparison is likely to attract a lot of attention (as it has for most past experiments) not only because of the political importance of the labor-effort effect but also simply because it is quantitative. “What is the labor-effort response in the experiment?” “It is X%.” That simple answer in the form of a nice, clear number that took years of research to produce can be very satisfying especially to an audience that doesn’t understand how far removed the raw comparison of control and experimental groups is from a prediction of the national labor-effort response to a fully-implemented UBI system.

Even as a measure of the initial response of workers, this comparison is likely to overestimate the effect of national UBI, because as explained above, the sample will probably be drawn from a small portion of the population, low-income workers—the people most likely to reduce their labor hours in response to UBI. Most of the U.S. NIT experiments drew entirely from this group. High-income workers who were likely to be unaffected were not tested. People living entirely off benefits and who might work more because UBI freed them from the poverty trap were also not tested. If future tests involve a similar focus, they will have to bring in a great deal of other evidence to make experimental evidence relevant to the overall question about labor effort.

A saturation study—depending on how economically diverse the site is—might include a wider range of income levels. Researchers will have to use evidence from other sources to estimate how the rest of the population will change their behavior. Economic theory and past evidence predicts that people with high incomes will not change their behavior very much (even in response to the higher marginal tax rates that UBI would require) and that people living on conditional benefits will work more on average.

The initial reaction of laborers is not the full effect on labor effort even in the short term. As earlier chapters explained, this effect depends on the reaction of employers to changes in laborers’ behavior and then on laborers response to that change, and so on. Basic supply and demand theory predicts that the market will react to a decline in labor hours by increasing wages and/or improving working conditions in the relevant sectors and that each of these effects will cause labor hours to rebound, partially counteracting the initial decline. RCTs cannot directly observe the labor-demand response at all, although they can use a microsimulation model to get some estimate of it. As always, that means that the experimental findings play a lesser role in determining the estimate—much of it coming from the assumptions going into the model. Saturation studies can capture some demand response, but only at the local level, which is likely to be much smaller than the national demand response.

Even these simulations will produce incomplete results because they will measure only the short-term response to a temporary program. The long-term response cannot as easily be estimated with simulation techniques, because it depends on unpredictable cumulative responses of variables, such as improved health, education, housing, cultural norms, food security, and so on.

But the simulations need to be run, and any unmeasurable long-term effects explained and perhaps predicted on an ad hoc qualitative basis, because of the central role that the labor-effort plays in the UBI discussion. A simple report of the raw comparison of the control and experimental groups, which could easily be three times the national response,⁴⁹ is likely to be misunderstood, sensationalized, or spun as a straightforward representation of the national response—greatly

biasing the discussion against UBI. This issue is crucial because an exaggerated understanding of the labor-market response dominated discussion of the NIT experiments in the 1970s.⁵⁰

The labor-effort claim is not merely that UBI reduces labor hours; it is the claim that the fall in labor effort is “unacceptably high.” The definition of unacceptable is subjective and morally loaded. UBI supporters are likely to define “acceptability” synonymously with sustainability, connecting it with cost (see discussion below). At least some opponents are likely to define it so strictly that they can present *any* decline in labor-effort as unacceptable. This extreme view was taken for granted in much of the 1970s discussion of the NIT experiments. Many writers considered the existence of any decline in labor hours as proof that BIG failed the test—regardless of how large or small that decline was and regardless of all other factors.⁵¹

Although in less wealthy nations, UBI has been associated with an increase in labor hours, in developed countries BIG is associated with a decline in labor hours. This might not be the result of the Finnish experiment, which it focuses *only* on people who are currently experiencing a poverty trap under the existing conditional system. Unless a nation has a very large number of people caught in a poverty trap or in extreme poverty such as that experienced in poorer nations, the average labor-effort response in a broad-based experiment is likely to be a slight decline.

That is, experiments will find a decline in the range that is “acceptable” by the sustainability standard and “unacceptable” by the no-decline-is-acceptable standard, giving each side the opportunity to spin the results their way. Researchers can head off the spin opportunity by recognizing it exists and addressing it directly. They can discuss the relevance of the experimental results to people with each of these points of view and look for other standards that might be of interest to people with more moderate views—in between sustainability and no decline at all. These standards might involve asking other questions, such as: how much of the decline was composed by workers reducing their hours, by unemployed workers increasing their search time, and by people leaving the labor force? If people do leave the labor force, how are they spending their time, as fulltime caregivers, as students, as entrepreneurs, and so on? What costs and benefits are associated with this decline in average labor effort?

The affordability claim and the cost issue in general

Experimental evidence plays a small but important role in estimating UBI’s cost, which is essential to **the affordability claim**. Most of the tax cost of UBI can be calculated with income statistics. That is, determine how much UBI costs assuming no one changes their behavior in response to it or to the tax increases that will accompany it. One recent estimate using that assumption shows that the United States could introduce a UBI large enough to eliminate poverty at a net cost of about \$539 billion per year.⁵² The role of experiments is to help determine how changes in people’s behavior affect that cost. Any negative labor-effort effect will increase that cost. Any increase in wages will decrease that cost. And any decline in the need for other social services (or any programs that can be replaced by UBI) will decrease that cost. Researchers can use experimental data on labor effort in simulation models to estimate these three effects.

That estimate is only for the tax-cost of UBI. Calculating the efficiency cost requires determining what portion of the labor-effort effect is attributable to the marginal tax rate alone. This again involves introducing other evidence and using statistical modeling, but it is important to compare the efficiency of UBI compared to other programs.

The complexity of the market and social reaction to UBI creates significant problems for reporting the findings of experiments. The affordability of UBI must be understood in the context

of the overall, long-term effects of a national program, but experiments can directly observe only one element in the long chain of reactions that determine those final effects—the short-term response of laborers in isolation from the response of employers. The rest—if it is not ignored—will have to come from microsimulations using data from other sources. Doing so means that the findings will be largely driven by the assumptions of that microsimulation model. Although the limitations of the experiments’ role in such a model is obvious to researchers, nonspecialists will need a great deal of explanation to understand it.

Estimates of the likely response of laborers to nonmarket income also exist in the economics literature. One could use those other-source estimates in a microsimulation to predict the market response to a given UBI with no experimental data. Experimental findings need to be understood as an effort to improve estimates of some of the parameters that go into the model necessary to estimate the variables of most interest to the UBI discussion.

Researchers in the 1970s tended to treat these limitations as if they were so obvious that they could be covered by a caveat or ignored completely, but they weren’t obvious to most people involved in the public discussion, and apparently few nonspecialists writing about experiments understood these limitations. Researchers can reduce that problem this time by clearly and explicitly explaining the limited role experimental data can contribute to our understanding of the cost of UBI. That role might well be extremely valuable despite its limits, but overblown beliefs about the role experiments can play needs to be headed off by emphasizing those limits.

The affordability claim differs from the cost question only by adding a judgment: how much is too much? Addressing it requires defining an affordability criterion. Unfortunately, the affordability criterion is subjective and at least partially morally loaded. UBI supporters (and perhaps others who are positively inclined toward UBI) are likely to define the affordability criterion synonymously with sustainability. That is, a program is unaffordable only if costs associated with it are so large that they collapse the program itself. Opponents are likely to define that criterion in such a way that any added cost is “unaffordable.” People in the middle might all have different criteria in mind. And many people probably have not carefully thought out the issue to settle on a criterion.

Researchers conducting experiments cannot hope to resolve the dispute or impose anything as the criteria. But they can examine questions that are relevant to the different ways that people who are interested in the UBI discussion view affordability. These might include: is a poverty-level UBI sustainable? What is the highest-sustainable UBI level? How much will UBI’s labor market and welfare effects increase or decrease its overall cost? What is the efficiency cost of UBI? How do the tax and efficiency cost of UBI compare to the cost of other programs capable of achieving similar goals? What affordability criteria are relevant in the local discussion of UBI?

The cost of some substantial level of UBI is almost certainly going to be sustainable but non-zero. That is, it will fall into a range where supporters can declare it “affordable” and opponents can declare it “unaffordable.” That fact makes any report on the cost of UBI vulnerable to spin from any side. Researchers and anyone else writing about the experiments can help head off that spin by recognizing the controversy over the affordability criteria. For example, they can report that the cost of this UBI scheme is affordable by these controversial criteria; unaffordable by these other equally controversial criteria. They can also consider how UBI compares in affordability to other programs of similar size and/or effectiveness. In other words, the affordability question needs to be connected to the cost-effectiveness question.

The economic-equality claim

The economic-equality claim, as stated, needs no test, because UBI necessarily reduces inequality through direct redistribution as long as it is set at a sustainable level. But the mere fact that UBI reduces economic inequality by some amount is not an important issue. The important issue is *how much* UBI reduces inequality. This question is partially testable because it depends on many market factors, some of which are observable. But experiments will only reveal the first step in a long chain of reactions that will determine UBI's effect on economic equality. Experiments can compare the incomes of people in the control and experimental groups, but they will need to combine that with evidence from other sources for UBI's like effect on taxes paid by higher-income people and on employers' wage response. Some kind of simulation will be necessary, and this estimate will be only the short-term effect of a temporary policy.

To get some idea of longer-term effects, researchers can observe the initial effects of UBI on education, health, safety, food security, and other factors that are correlated with economic mobility, but they cannot actually observe whether those factors do lead to greater economic equality for experimental participants. Researchers can use other evidence about how these variables are correlated to economic mobility to estimate their effect on economic equality. Actual experimental findings will make only a small contribution to that estimate, and because they will also need outside evidence to estimate the full impact of UBI on those contributing factors, the effort becomes somewhat speculative.

The widespread benefit claim

The widespread benefit claim, as I use the term, is distinct from the question of harm and benefit to workers. It is the claim that UBI's benefits—whatever they might be—are widely shared by many people at any given time and by a significantly greater portion of people will benefit from UBI at some time in their lives. The question of how many people can expect to benefit at some time during the course of their lives is clearly as important as the question of how many people benefit at any given time. Also, it frames the issue differently. Simply counting contributors and beneficiaries can give the impression that those categories are fixed classes, when in fact many people move back and forth between the two groups during the course of their lives.

The spread of UBI's benefits is determined largely by its structure. Assuming no changes in behavior, a reasonably "affordable" UBI (about 3% of GDP for the United States) can be structured so that 40% of the population receives some net financial benefit from UBI at any one time. That is, the additional taxes they pay to support the UBI are smaller than the UBI they receive. A UBI costing 10% could be structured so that well over 60% of people benefit in this sense.⁵³

This much is sufficient to say that a large portion of the population benefits at any one time. Researchers will have to use evidence about economic mobility to determine how many more people can expect to benefit financially at some time in their lives. That is, everyone whose income goes below the breakeven point at some point in time between birth and death.

One argument for UBI supposes that it will have a positive psychological impact on almost everyone's welfare regardless of income both through community effects and through permanently removing the fear of poverty and destitution. Although no one should assume that everyone will benefit all-things-considered, there is evidence that more equal societies are in many ways better

for everyone.⁵⁴ Unfortunately, RCTs are unable to provide any direct evidence about the community or psychological impact on net (financial) contributors. A saturation study can use surveys to attempt to gauge whether net contributors feel more secure living in a community without poverty but will be biased in two opposite directions, because the impact of taxes and national-level community effects will both be unobserved.

Direct observation of the widespread-benefit claim would require an extremely long-term study involving subjects at all levels of income. Experiments can make two small contributions toward understanding the widespread benefits claim by observing the labor-effort effect and the impact on welfare factors known to be correlated with economic mobility. The labor-effort effect is the first step in a chain of possible effects that might lead to greater benefit at any one time and over time. Improvements in housing, food security, safety, health education and so on are the first step in a chain of possible effects that might lead to greater economic mobility over time.

Again, UBI experiments can only contribute a small piece of evidence to the effort to make these estimates. Computer simulations and other evidence will again be necessary, and their ability will be limited as well. But a focus on how people benefit throughout their lives is essential to a good public understanding of UBI's likely effects.

The harm-to-workers claim and the benefit-to-workers claims

The harm-to-workers claim and **the benefit-to-workers claim** are both things that experiments can say *something* about, but researchers need to approach them cautiously because what experiments can say is very limited, easily misinterpreted, and connected to contentious ethical disagreements, such as the exploitation debate (discussed in Chapter 15). But they cannot be ignored because many of the variables that experiments address will appear to say much more about these claims than they actually do. Researchers have to address these claims to avoid allowing misunderstanding and spin to create an exaggerated impression of the effects of UBI on these issues.

The harm- and benefit-to-workers claims present at least two difficult subjective definitional issues.

First, what do we mean by “workers?” Is a fulltime parent or caregiver a worker? Are other unpaid workers “workers?” Is a person living off financial investments a worker? How many hours per week does a part-time laborer have to be employed to count as a worker? How many weeks can an unemployed person be unable to find a job and still count as a worker? Is a person who uses UBI for a one-year sabbatical from a 40-year working life a worker? Do children, the elderly, and the disabled count as “workers,” and if not, do we judge any harm that these groups create for workers differently than harm to workers created by other non-working people? And so on. This ambiguity is why most of this book avoids the term “worker” altogether in favor of the clearer term laborer (meaning a person working for pay). But this section uses it because the ambiguous idea is what matters for the debate over this pair of claims.

Second, what do we mean by harm and benefit? Financial harm and benefit are relatively easy to observe and far more quantifiable than overall benefit, but they are clearly less important. And so, it is best to consider both.

One thing is certain: both of these claims—as stated—are oversimplified. Almost any UBI system will benefit some workers and harm others at least in a financial sense. One can imagine community effects so strong that every worker ends up either better or worse off overall, but no such scenario appears plausible. The relevant questions seem to be: which workers benefit and

how much? Which workers are harmed and how much? Is there evidence that a group of people will abandon all “work” (however defined); if so, how many will, and how much will this group cost workers? But it must be remembered that not everyone agrees that the existence of such a group is ethically wrong. Research has to be presented in a way that doesn’t lead people to presume one controversial moral position or another.

As with the affordability claim, experimental evidence plays only a small role in calculating the harm and benefit to workers. Most of the financial harm and benefit of a UBI system is determined by its structure and does not need a test. If UBI is largely income-tax financed, anyone who makes less than the break-even point benefits and anyone who makes more is financially harmed assuming. Other ways of financing UBI make calculating the breakeven point more difficult, but all financing methods create winners and losers. As with costs, we get a good idea of what they are from the structure of the program, counterfactually assuming that nobody changes their behavior in response to the UBI.

Because a large UBI system can be structured to financially benefit the lower 40-60% of the population even without community effects,⁵⁵ a UBI system will definitely benefit a very large number of workers. Only workers in the upper 40-60% of the income distribution might be financially harmed; many net contributors will also be nonworkers; and the direct financial harm to workers in the low end of the net-contributory range is likely to be small and possibly overridden by community effects.

Researchers can help avoid misunderstanding by presenting findings for various demographic groups and various definitions of workers. What percentage of workers are financially harmed? What percentage are financially helped? What is the average net benefit to the average net-beneficiary worker? What is the average net harm to the average net-contributory worker? What are the average before-and-after-tax-and-transfer incomes to the average net-beneficiary worker and the average net-contributor worker? What percentage of UBI net benefits go to other people in other demographic categories of interest to the discussion, who might not be defined as “workers.” These might include children, caregivers, retirees, students, and so on.

Researchers will understandably reject making the controversial judgment of identifying a group of people as those who could work, should work, and don’t work under UBI. To anyone who feels strongly about non-wealthy people’s supposed responsibility to labor will be more concerned with this number than anything else. People with opposite feelings (such as those who endorse the freedom and productive non-labor claims) will also be interested in some number like this.

Because these issues are so important to the political economy of UBI, and because so many experimental results can be overblown to implying much more about these issues than they actually do, researchers can greatly improve the likelihood of raising the level of debate by trying to find some non-judgmental way to report numbers that usefully inform people who are concerned with these issues. One way to do it might be to report the percentage of the cost caused by the benefits to people in the various demographic categories relevant to the national debate, if there is some way to do so without inviting judgment and without communicating the perception that all net contributors are workers.

Similarly, the labor-effort effect—the principal potential source of harm to workers—is also a direct benefit to workers, depending on how both of those terms are defined. If the new experiments find similar results as the 1970s experiments, most of the decline in labor effort will not be explained by people withdrawing from the labor market, but by people working fewer hours and by people who happen to become unemployed taking longer to find their next job. Both of

these are *benefits* to workers unless someone takes a restricted definition of worker to include only people who spend the maximum possible time in paid labor. Furthermore, the knowledge that workers don't have to take the first job they find if they happen to become unemployed is a benefit to all workers. Reporting on these benefits to workers and the associated costs is also useful to avoid misunderstanding.

The contribution of UBI experiments to these issues is to determine how changes in the behavior and health of net-recipients affect the relevant numbers above, and as before, what they can say about it is limited but still significant.

Most of the experimental contribution to the understanding of financial harm to net-contributory workers is determined how large the total cost of UBI is. Policymakers can choose to spread that burden in many different ways, some of which would put most of the burden on rent-paying assets rather than on labor income. This difference will have different moral implications with people who have different ethical positions. Tests will say nothing about how policymakers will choose to spread the burden among net contributors.

Experiments can say much more about the benefit to workers. Workers working less is the first step *both* in the story ending in worker harm *and* in the story ending in worker benefit. The harm-to-workers claim relies on reduced labor effort increasing the net cost of the program and creating the need for higher taxes. The benefit-to-workers claim relies initially on reduced labor effort followed by improved labor market conditions increasing pay and improving working conditions. Increased pay—even if it is just among net-recipient workers—would mitigate some of the increased tax cost caused by the initial decline in work effort and it would also mitigate some of the decline in work hours, again decreasing cost. The reason employers respond to a drop in labor supply by increasing wages and improving working conditions is to entice some of those workers to put in more hours.

RCTs are completely incapable of examining any steps in these chains of reaction beyond the first, and they can only partially and indirectly examine the first, as discussed above. A saturation study can shed some light on subsequent reactions. But to extrapolate to national results researchers would need to bring in data from other sources on the elasticity of supply and demand in various labor markets to estimate the effects of a simulation model. And of course, the outcome of the model will be driven by a great deal of nonexperimental evidence.

Trials will be able to give the most conclusive answer if the labor-market response is small. If so, the financial harm and benefit to workers will be pretty much dictated by the structure of the program. But if the labor market changes are large, it becomes difficult to tell whether those changes increase or decrease the benefit to workers. This unfortunate fact doesn't take away from the value of the data. The studies might reveal which segments of the labor market (in terms of occupation, income level, etc.) will be most affected.

Trials will also contribute to the understanding of the costs and benefits to workers through possible reductions in social costs and through possibly improved worker productivity. These issues are discussed above. Once again, experiments will only observe part of the first step in a long chain that might lead to substantially increased benefit and decreased costs for net-contributory and net-beneficiary workers alike. The rest will have to be filled in by nonexperimental evidence to come up with the estimate.

The cost-effectiveness claim

Chapter 12 argued that **the cost-effectiveness claim** should be the bottom line question. It is probably the most important empirical question that any researcher can address about UBI. But it requires little additional discussion because it can be examined by putting together the evidence discussed above. Each variable discussed above can be looked at individually in cost-effectiveness terms, and all the variables of interest can be indexed into one overall cost-effectiveness estimate. Chapter 12 also argued that the cost-effectiveness question is best formulated in relative terms—relative to other policies of a similar size aimed at achieving similar goals, but doing so requires either gathering more experimental data from a second experimental group or bringing in more nonexperimental data to estimate the effects of the policy UBI is to be tested against. Neither of these options are attractive, but they're necessary to answer the questions that most need answering.

Combining experimental and theoretical information into an answer to the cost-effectiveness question requires adding more theory, making the results one step less direct and conclusive, requiring one more discussion of limits of experimental findings, but it is more important to report less conclusive answers to meaningful questions than more conclusive answers to less meaningful or misleading questions.

Chapter 17: From the dream test to good tests within feasible budgets

Putting the last three chapters reveals one important fact: after you eliminate the claims that don't need a test, the claims that can't be tested, and the claims that can only be tested partially, indirectly, or inconclusively, no claims are left from the list in Chapter 12. I have been unable to find any relevant claims that can be tested fully, directly, and reasonably conclusively in a small-scale experimental setting the way the individual effectiveness of a medicine can often be tested. Experimenters have to do their best to enlighten our understanding of those difficult-or-impossible-to-test claims.

In light of this issue and all the difficulties with existing tests discussed above, this chapter starts with the dream test and works down toward more feasible tests. The “dream test” is one in which money, time, and political will present no obstacles to testing virtually all the empirical claims that matter in the UBI discussion. Although the dream test is conceivably within human control, it would take an act of united will large enough to make it effectively impossible, but by considering it, we can learn something about the daunting task of trying to answer the most important questions about UBI.

Imagine all the nations of the Earth are democratic and all of them share an overriding desire to test UBI at the national level. They are willing to devote as much time and resources as the test needs. They could divide the 200 or so nations of the world into control and experimental groups, effectively combining RCT and saturation study techniques with enough saturation sights to ensure statistical significance for most variables of interest. Researchers could then run the test for 50 or 100 years—extending it as long as it takes until the long-term effects seem to have fully played out.

Human beings are capable of running this test if enough of them want to, but it is a very strange thing for humans to want to do. The early generations would have to be more interested in testing UBI at the national level than they are in the question of whether or not to introduce UBI

in their nation during their lifetime. And they would have to maintain politically united in that desire for a century. It's not going to happen, but this test is the only kind capable of solving most of the problems discussed throughout this book.

All those national-level community effects and all those long-term effects that obscured the relevance of experimental data for every claim considered above would no longer be obstacles. Most of the variables the identified above would become testable in a statistically useful way. Some problems would remain, such as observational difficulties and the problem that it is much easier to produce definitive answers for quantitative questions than for qualitative questions. But the observable, empirical differences between the nations with UBIs and those with other policies would become so apparent that it is hard to believe anyone would still be unable to make up their mind. Pretty much, the only remaining disagreements would be entirely ethical in character, and we have a good chance of learning so much that ethical positions might converge for a substantial majority of people.

This utterly infeasible experiment is the only kind of experiment capable of directly representative results for any variable of interest. The experiment might be able to make do with less than all the world's countries in less than a half a century, but we don't work down from this test to a feasible-sized experiment without losing the ability to observe many or perhaps most of the community and long-term effects that the UBI discussion hinges on. But yet if we simply drop the difficult-to-test variables from consideration, we do the scientific equivalent of looking under the streetlight.

Therefore, perhaps, the second-best test of a UBI is to introduce it in a single country. This test would sacrifice the ability to *control* for anything, but it maintains the ability to observe all the relevant effects of UBI and is an excellent opportunity for trial and error. Is it better to make controlled observations of a few of UBI's effects or uncontrolled observations of all the things we really want to know about UBI? The answer depends on the importance of the effects lost in the controlled experiment and on what other techniques are available to account for the lack of control in the national trial.

Iceland is a country of only 335,000 people, and it would make an excellent national saturation site. Bermuda is even smaller, only 65,000 people, almost to the point where one could imagine the wealthy institutions paying Bermuda to introduce a full national UBI on an experimental basis. GiveDirectly's study in Kenya includes 16,000 people, making it larger than some of the world's smallest countries. But—as an experiment—a nationwide study is highly unlikely. The most-promising possibility along these lines is that some country might decide on its own to introduce UBI and provide a natural experiment. Waiting for this to happen has obvious drawbacks as a research strategy.

The next best (and probably still too expensive) study could combine RCT and saturation techniques. The experimental group and control group would each need to be comprised of two-dozen or more communities to control for unobserved differences between sites, and it would be best to have two-dozen more additional communities receiving a more generous version of existing programs—or whatever program UBI is being tested against. I've already discussed the difficulty of testing UBI against an equal-sized alternative policy, so the rest of the discussion leaves it out, even though it would greatly increase the value of the data collected.

Ideally, the saturation sites would be selected to be demographically representative of the nation as a whole in as many ways as possible with maybe one or two exceptions. Isolated communities might be preferred despite introducing statistical bias because, as Chapter 4 mentioned, they are more likely to show the kind of community effects we can expect at the

national level. Researchers might want to focus on poorer communities because those are the ones where UBI will have the most important impact, but doing so makes the results very different than national results, creating the need to extrapolate from other sources to get national estimates.

As large as the integrated RCT and saturation study would be, it nevertheless loses the ability to estimate many important community effects of UBI, because for so many variables of interest to the issue, the relevant community size is the nation as a whole. A relatively isolated site will have community effects more like those at the national level, but almost certainly much smaller. Any attempt to estimate those effects will have to come from other sources. Yet, local community effects are important, especially in impoverished areas, and so a test like this is worth doing if feasible.

Studies fully integrating RCT and saturation techniques are likely to be prohibitively expensive in wealthier nations, but they are possible in relatively poor countries. The Indian experiment used multiple saturation sites. And the GiveDirectly study in Kenya is the first one with enough saturation sites to statistically control for unobserved differences between communities.

If multiple saturation sites are unaffordable, the next best experiment might be a combination of one saturation study and one individualized RCT. The saturation study would require two sites, one experimental group and one control. The word control is, in this case, a bit of a misnomer because a study like it cannot control for differences between the sites. For example, imagine that after the study began, the largest employer in the control site went out of business, causing a surge in unemployment in that community. A simple comparison of employment hours in the two communities would associate UBI with an increase rather than a decrease in labor effort. If something as dramatic as a major employer going out of business happens, researchers will probably observe and do their best to take account of it. Manually taking account of it is not easy or terribly precise, but the bigger issue to statisticians is unobserved causes. The very idea of an RCT is to use randomization to control for observed factors. A study involving only one experimental saturation site has no such ability.

Yet, a saturation study is valuable at this level. It provides uncontrolled observations of local community effects, while an RCT alone provides no observation at all of local community effects. Uncontrolled observations are better than none. By running one study of each type researchers can get both kinds of observations.

One way to increase the reliability of saturation study results is to begin observing the two communities a year or two before flipping a coin to see which one becomes the experimental site and which one becomes the control. This method effectively allows the site receiving UBI to be tested both against the control site and against itself before the introduction of UBI, and both comparisons can be done across the two sites. Researchers can then look for other factors that might cause the two communities diverge. There might always be unobserved factors that cause diverge between the two communities to begin at the same time as the UBI, but very often the likelihood is against such significant factors going unobserved.

If a saturation study is not possible, an individualized RCT will have to go it alone. The RCT is a good, scientific technique, but unfortunately, it is one that is not able to give direct answers to any of the questions relevant to the public discussion of UBI. This shortcoming does not make an RCT useless for studying UBI, but it does make it far more difficult to conduct and report the results of a UBI experiment in ways that truly inform the public discussion of UBI.

I suggest formulating research questions aimed toward empirical issues that are relevant to the nation's decision whether or not to introduce it. Chapters 12-16 have made an effort to identify

many of them, but no one should take my word for it. The discussion varies considerably country-to-country. As argued above, it is important to formulate questions relevant to the local discussion without taking shortcuts around the lessons here. It is better to concentrate on what people most want to know about UBI than on what they would most like to get out of an experiment. Framing the question toward what people want specifically from an experiment is like asking them to look under the streetlight.

And as this book has stressed throughout, once experimental results are in, a simple report on the differences between the control and experimental groups is highly inadequate. Someone has to extrapolate from those results to estimates of the things people really want to know, if any feasible test is going to succeed in raising the level of public discussion of UBI.

Whose job is it?

The last several chapters have discussed what “researchers” can do, but much of what I have suggested will be outside the field of specialization of many people involved in conducting and writing on UBI experiments. To say that “researchers” can do it is not to say that some particular researcher can or should or must do it. Nor is it to say that researchers are the most responsible party. Not all of it is necessarily the responsibility of or within the ability of all the researchers conducting UBI experiments. Their job is to do what they were hired to do. The efforts I suggest are not necessarily anyone’s job at this time. The people commissioning research have the most power to address these issues—even if they don’t have the most expertise.

The goal of this book is not to blame researchers or anyone else for the difficulties of testing UBI and making that test understood. The problems stem from the inherent difficulty of testing UBI and the long chain of connections and likely misunderstandings between the diverse people involved with UBI experiments and with public decision making about UBI. No person in that chain bears responsibility. I just hope the suggestions here are useful to everyone involved and everyone watching.

Notes

¹ UBI can, of course, be a regional policy. The rest of the book let’s that go without saying.

² The Basic Income Earth Network defined UBI this way at its 2016 meeting in an effort to reflect common usage.

³ {Standing, 2012 #1357}

⁴ {Standing, 2012 #1357}

⁵ {Wilkinson, 2009 #947}

⁶ {McCambridge, 2014 #1414}

⁷ Thanks to Evelyn Forget for alerting me to this last issue.

⁸ {Robins, 1984 #1413}; {Widerquist, 2005 #208}

⁹ {Atkinson, 1995 #170}; {Widerquist, 2017 #1408}

¹⁰ {Widerquist, 2017 #1408}

¹¹ {Atkinson, 1995 #170}

¹² {Widerquist, 2017 #1408}
¹³ See Chapter 6.
¹⁴ {Moffitt, 1979 #1402}. See Chapter 6 for further discussion
¹⁵ {Widerquist, 2005 #208}
¹⁶ {Burtless, 1986 #1400} {Keeley, 1981 #1401} {Robins, 1985 #1399}
¹⁷ {Moffitt, 1979 #1402}
¹⁸ {Greenberg, 1983 #1404}
¹⁹ {Bishop, 1979 #1405}
²⁰ {Levine, 2005 #210}
²¹ {Widerquist, 2005 #208}
²² {Levine, 2005 #210}
²³ {Levine, 2005 #210}; {Widerquist, 2005 #208}
²⁴ {Widerquist, 2005 #208}
²⁵ {Widerquist, 2005 #208}
²⁶ {Widerquist, 2005 #208}
²⁷ {Widerquist, 2005 #208}
²⁸ {Burtless, 1986 #1400}
²⁹ {Widerquist, 2005 #208}
³⁰ {Forget, 2011 #978}
³¹ {Haarmann, 2009 #989}
³² {Standing, 2013 #1409}
³³ {Condliffe, 2016 #1406}
³⁴ {Widerquist, 2012 #849}; {Widerquist, 2012 #850}; {Hanlon, 2010 #1002}; {Standing, 2008 #979}
³⁵ {Miller, 2017 #1411; Standing, 2017 #1410; Van Parijs, 2017 #1391; Widerquist, 2013 #995; Torry, 2016 #1412}
³⁶ {Widerquist, 2017 #1408}
³⁷ However, they might not share the same vision of UBI. Therefore, similarity in motive doesn't imply that they will test the same version of UBI that supporters are most interested in. A UBI test cannot be as diverse as the UBI movement is.
³⁸ See Chapter 6.
³⁹ {Widerquist, 2006 #1407}
⁴⁰ {Widerquist, 2006 #1407}
⁴¹ I use "popular writers" to mean people who write for nonspecialists (the populace), not to mean people who have a lot of readers.
⁴² {Widerquist, 2006 #1407}
⁴³ {Widerquist, 2005 #208}; see also Chapter 6 above.
⁴⁴ See Chapter 6 for how this happened for the labor market findings of the 1970s experiments.
⁴⁵ Similarly, people with differing ethical beliefs might give a higher moral priority to a less efficient system that forced non-wealthy people to accept employment than to a more efficient redistribution system that gave them the opportunity to refuse employment.
⁴⁶ {Widerquist, 2006 #1407}
⁴⁷ See subsequent chapters.
⁴⁸ See chapter 6 and {Widerquist, 2005 #208}
⁴⁹ {Moffitt, 1979 #1402}
⁵⁰ {Widerquist, 2005 #208}

⁵¹ {Widerquist, 2005 #208} Also see Chapter 6.

⁵² {Widerquist, 2017 #1408}

⁵³ {Widerquist, 2017 #1408}

⁵⁴ {Wilkinson, 2009 #947}

⁵⁵ {Widerquist, 2017 #1408}. See discussion earlier in this chapter.