Karl Widerquist

The Devil's in the Caveats: A Brief Discussion of the Difficulties of Basic Income Experiments

Karl Widerquist
Georgetown University-Qatar

The devil is in the detail is a common saying about policy *proposals*. Perhaps we need a similar saying about policy *research*, something like *the devil is in the caveats*. No simple list of caveats can bridge the enormous gap in understanding between the specialists who conduct policy research and the citizens and policymakers (including both elected officials and public servants) who are responsible for policy in a democracy, but who often have overblown expectations about what policy research can do.

For example, the headline, 'In 2017, We Will Find Out If a Basic Income Makes Sense', which appeared in *MIT Technology Review* in December 2016 (Condliffe 2016), expressed a common belief about experiments with the Universal Basic Income (UBI) – a policy to put a floor under everyone's income. Although the most laughable inaccuracy of this headline was there were no plans to release findings from any experiments at all in 2017 (nor were any published), the more important inaccuracy was that it reflected the common, but naive belief that UBI experiments are capable of determining whether UBI 'makes sense'. No social science experiment can do any such thing. Social science experiments can produce valuable information, but they cannot answer the big questions that most interest policymakers and voters, such as does UBI work or should we introduce it?

The limited contribution that social science experiments can make to big policy questions like these would not be a problem if everyone understood the experiment's limitations, but unfortunately, the article in *MIT Technology Review* is no anomaly. It is a good example of the misreporting on UBI and related experiments that has gone on for decades (Widerquist 2005) by the publications we count on to get it right. *MIT Technology Review* was founded at the Massachusetts Institute of Technology in 1899. Its website promises "intelligent, lucid, and authoritative […] journalism […] by a knowledgeable editorial staff, governed by a policy of accuracy and independence".[1] Although *the Review's* expertise is in technology rather than scientific research, it is the kind of publication that one would expect to be most able to help non-specialists understand the limits and usefulness of scientific research.

Although there is some overlap between the academics, journalists, policymakers, and citizens involved in policy research and policy discussions, most of the individuals in these groups do not have enough shared background knowledge to understand each other well. Researchers often do not understand what citizens and policymakers expect from research, while citizens and policymakers frequently fail to grasp the inherent difficulties of policy research, or the difference between what research shows and what they most want to know. People who do not understand the limits of experiments also cannot comprehend the value that experiments can contribute to our understanding of an issue.

Specialists usually include a list of caveats covering the limitations of their research, but caveats are incapable of doing the work researchers often rely on them to do. A dense, dull, and lengthy list of caveats cannot provide non-specialists with a firm grasp of what research does and does not imply about the policy at issue. As a result, even the best scientific policy research can leave non-specialists with an oversimplified, or simply wrong impression of its implications for policy. Better written, longer, clearer caveats will not solve the problem either. The communication problem, coupled with the inherent limitations of social science experimentation, call for an entirely different approach to bridging the gap in understanding.

My forthcoming book, *A Critical Analysis of Basic Income Experiments for Researchers, Policymakers, and Citizens*: *The Devil's in the Caveats*, addresses how these sorts of problems affect Universal Basic Income (UBI) experiments that are underway, planned, or being considered in several countries around the world at present. This article previews and summarises the major findings of that book.

UBI has many complex economic, political, social, and cultural effects that cannot be observed in any small-scale, controlled experiment. Therefore, even the best UBI experiment makes only a small contribution to the body of knowledge on the policy in question and leaves many important questions unanswered. Citizens and policymakers considering introducing UBI are understandably interested in larger issues. They want answers to the big questions like does UBI work as intended; is it cost-effective; should we introduce it on a national level? The gap between what an experiment can show and the answers to these big questions is enormous. Within one field, specialists can often achieve a mutual understanding of this gap with no more than a simple list of caveats, many of which can go without mentioning. Across different fields mutual understanding quickly gets more difficult, and it becomes extremely difficult between groups as diverse as the people involved in the discussion of UBI and UBI experiments.

[1] MIT Technology Review, What We Do, https://www.technologyreview.com/about/.

The process that brought about the experiments in most countries is not likely to produce research focused on bridging that gap in understanding. The demand for the current round of experiments seems to be driven more by the desire to have a UBI experiment than by the desire to learn anything specific about UBI from an experiment. An unfocused demand for a test puts researchers in a position to learn whatever an experiment can show, regardless of whether it is closely connected to what citizens and policymakers most want to know. The vast majority of research specialists who conduct experiments are not fools or fakers. They will look for evidence that makes a positive and useful contribution to the body of knowledge about UBI. But the effort to translate that contribution into a better public understanding of the body of evidence about UBI is far more difficult than often recognised. This communications problem has badly affected many past experiments and is in danger of happening again.

To understand the difficulty of the task, imagine a puzzle strewn out over the floor of a large, dark, locked room. A map of the entire puzzle, assembled together, provides an answer to the big questions – does it work, and should we implement it. An experiment shines a light through a window, lighting up some of the puzzle pieces, so that researchers can attempt to map how they might fit together. They can easily map the pieces near the window, but further away their view gets dimmer, the accuracy of their map decreases, and in dark corners of the room, many pieces remain entirely unobserved. Although scientists like to solve entire puzzles when possible, under normal circumstances, they have to settle for something less ambitious. That is why the basic goal of scientific research is to increase the sum of knowledge available to the scientific community – even if that increase is very small. In terms of the example, if a research project can map even one new piece of the puzzle, it succeeds in the basic goal, even if the puzzle as a whole remains unsolved and the map is only readable to other scientists.

As the headline mentioned above illustrates, non-specialists tend to expect something far more definitive from social science experiments, often assuming they have the same goal as high school science tests: to determine whether the subject passes or fails. People often expect that experimental researchers will produce an estimate of whether UBI works or whether the country should introduce it. In terms of the metaphor, they expect researchers to solve the entire puzzle; or at least to provide their best estimate of that solution. If researchers present their findings as social scientists normally do, they present something fundamentally different from what citizens and policymakers are looking for and possibly expecting. The potential for misunderstanding is enormous when research reports say something to the effect of *here are the parts of the puzzle we were able to map* to an audience looking for something to the effect of *here is our best estimate of the solution to the entire puzzle.* Caveats do not and cannot draw the necessary connection: *here is how the parts we were able to map can be used toward a larger effort to find the solution to the entire puzzle and how close or far we remain from it.*

Caveats tend to focus, not on the connection between the two goals, but on trying to help people understand the research on its own terms. In terms of the analogy, caveats tend to focus on the areas that experiments were able to map: how did they map this area; what does it mean to map this area; how accurate is the map of this area, and so on. The relationship between the areas mapped and the solution to the whole puzzle is often covered by one big caveat so seemingly simple that it often goes unstated: obviously the areas we mapped are far from a solution to the entire puzzle. In other words, the information gathered about UBI in an experiment is far from a definitive, overall evaluation of UBI as a policy. As obvious as that caveat might be to researchers, it is not at all obvious to many non-specialists.

Of course, non-specialists know there are some caveats about the reliability of the experiment, but if they overlook or misunderstand that one big caveat they will nevertheless believe that researchers provide their best estimate of whether 'Basic Income Makes Sense' (Condliffe 2016) and they will tend to look for that answer in any report on the study. If they get no help doing it, they are likely to overestimate the political implications of the information that experiments find, providing a great opportunity for spin and sensationalism by people willing to seize on small findings that sound positive or negative as proof that the programme has been proven to be a success or a failure. The book and some of my previous work argue that earlier UBI-related experiments have been misunderstood and misused in these ways (Widerquist 2005).

The difficulties above follow on from the complexity of the science involved. Now consider how ethics further complicates the issue. In terms of the analogy, this puzzle is a very special kind: the pieces fit together in different ways depending on one's moral values. If research definitively proves that a policy does not achieve the goals that its supporters hope it does, research can give a conclusive answer without dealing with ethical controversy. But if a sustainable policy achieves some goal and has some side effects, reasonable people can disagree about how good or bad those goals and side effects are; and how we should evaluate trade-offs between them. Therefore, reasonable people can disagree about whether the evidence indicates that the policy works and should be introduced; or whether that same evidence indicates the policy does not work and should not be introduced. This problem greatly affects the UBI discussion because supporters

and opponents tend to take very different moral positions.

Many people, including many specialists, are less than fully aware of the extent to which their beliefs on policy issues are driven by empirical evidence about a policy's effects; or by a controversial moral evaluation of those effects. For example, mainstream economic methodology incorporates a great deal of utilitarianism, which was the prevailing ethical framework when basic mainstream economic techniques were developed, but has long since lost its prominence in political philosophy and political theory. Many articles in economics journals read as if the author is unaware of the moral judgments incorporated into that methodology. Additionally, not everyone is honest about the extent to which their policy judgments are driven by controversial moral judgments. Some will try to spin the results by hiding the extent to which their evaluation of the evidence is driven by their moral position and portray it as the only objective reality.

Into this ethical morass falls the dense and difficult research report of an experiment's findings with an often tedious and easily ignorable list of caveats about the research's limitations and usually a complete absence of discussion about the moral judgments needed to evaluate the study's implications for policy. Under such circumstances, no one should be surprised that social science experiments easily fall victim to misunderstanding, spin, sensationalism, and oversimplification. Perhaps we should expect these problems to happen more often than not. After all, it is easier to understand an oversimplification than genuine complexity.

Solutions to these problems are difficult and imperfect, but we have to try to address them, if UBI experiments are going to achieve their goal. I presume the overall goal of UBI experiments is (and should be) to enlighten the public discussion by increasing public understanding of evidence on UBI. I do not think that this goal is controversial or new. And I will argue that it should be endorsed by virtually any UBI-related experiment no matter what other goals it might have, such as the basic goal of scientific research (mentioned above), working out technical issues that are important to policymakers, or in some cases, politically promoting UBI. There is nothing inherently wrong with using a study – even a small-scale, less-rigorous study – to promote a policy, as long as the evidence is presented honestly and aimed at improved understanding. In other words, the need to keep the goal of enlightening discussion through good communication and an orientation toward the most important issues is as important to the most political UBI demonstration project as it is to a more rigorous study.

Some past researchers (either conducting or writing about experiments) have failed to appreciate how difficult it is to accomplish this goal, especially when they focus primarily on the basic goal of scientific research. Increasing the amount of knowledge available to the scientific community does not necessarily or easily translate into improve public understanding of that evidence. The gap in background knowledge has to be addressed because it creates risks that less politically oriented research does not entail, including the vulnerability to misunderstanding, spin, misuse, sensationalism, or oversimplification.

Perhaps the main message of Widerquist (forthcoming 1) is that UBI experiments seldom if ever succeed in enlightening public discussion merely by trying to get non-specialists to understand experimental findings on their own terms. It is not enough to say, *here are the pieces of the puzzle we managed to map.* It is not enough to explain what experimental group is, what a control group is, and what the differences were between the two groups in the study. It is not enough to have a new and improved list of caveats about experimental limitations. Experimental findings should not be presented as a stand-alone piece of research, but as a small part of a larger effort to use all available evidence to answer the big questions about UBI; and to explain the extent to which the big questions remain unanswered. Researchers must attempt to find the information that will be of the most value to the public discussion, and someone – not necessarily the researchers conducting the study – has to attempt the difficult task of communicating those results in a way that people involved in the public discussion of the issue will understand. The difficulty of these tasks is at least half of what the book is about.

Widerquist (forthcoming 1) discusses the difficulty of conducting UBI experiments and communicating their results given both the inherent limits of experimental techniques and the many barriers that make it difficult for researchers, journalists, policymakers, citizens, and anyone else interested in UBI or UBI experiments to understand each other. The book's goals are to improve both the experiments and public understanding of them. Therefore, with the experiments' goal of enlightening public discussion in mind, this book asks two distinct but closely related questions:

1. How do you do a good experiment given the difficulties involved?
2. How can citizens, policymakers, researchers, journalists, and others interested in UBI and UBI experiments communicate in ways that will lead to a better public understanding of the implications of UBI experiments for the public discussion of UBI?

This project is an applied examination of a family of problems *specific* to UBI experiments with no claim that these problems are necessarily *unique* to UBI experiments. Many such difficulties apply to all social

science experiments, and some apply to all policy-related research – see also Deaton and Cartwright (2016); and Teele (2017). To the best of my knowledge, my book will be the first to focus entirely on applying this kind of analysis to UBI experiments, but will not explore whether the kinds of problems discussed for UBI experiments are as bad or worse than problems involved in other social science experiments.

This article and book are written for anyone interested in UBI experiments and UBI as a policy – they are for researchers, journalists, policymakers, citizens, and people who partly belong to one group and partly to another. The danger of misunderstandings exists between everyone involved; and everyone involved can help solve them. No single group can easily clear up such misunderstandings on their own; and hopefully we can all benefit from thinking through the problems examined by this book. Policymakers, journalists, and citizens who understand the place of experiments in the political economy of the UBI discussion will be able to communicate their desire for experiments that are more relevant to that discussion. They will learn more from any experiments that are conducted. And they will be better equipped to counter spin and sensationalism.

Researchers who understand the place of experiments in the political economy of the UBI discussion can obviously communicate their results more effectively. But it is not just about communication. Researchers who understand and respect the public discussion can design better experiments. It would be a mistake to believe researchers conducting experiments can resolve all of these communications issues on their own. Although research specialists are professionals at communicating with other specialists, the vast majority of them are amateurs at communicating with non-specialists – and I am no exception. Scientists are trained to conduct research and communicate it to other scientists, but have no special training in the skills needed to bridge the communications gap. Very often specialists do not know what evidence would be most valuable to citizens or policymakers or how best to help citizens and policymakers understand the value of the evidence researchers are able to find. The ultimate responsibility rests more with the policymakers and donors *commissioning* experiments than with the researchers conducting experiments. They – or whoever they put in charge of hiring the research to conduct experiments – might have the most to gain from understanding the communications gaps involved in UBI experiments.

As more experiments get underway and present their findings, it is important to consider lessons in how to improve the chances that experiments will successfully enlighten the public discussion of UBI. As the book argues, past UBI-related experiments – despite almost always being good science – have a mixed record at increasing the understanding of evidence among non-specialists. Some succeeded and some failed. No matter what the primary goals of an experiment are, the people commissioning and conducting them ignore the public role of UBI experiments at their peril. The primary goal of a UBI experiment, might simply be to examine a few narrow technical issues that are of particular interest to policymakers commissioning the study or to the research community. There is nothing wrong with the desire to make some goal like this the main focus of a project. But UBI experiments are too closely tied to the political process and their results are too easily misunderstood for researchers to ignore experiments' role in the political economy of the UBI discussion.

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 recent years. It is no coincidence that UBI-related experiments took place in the 1970s and not again (almost anywhere) until the 2010s. These efforts corresponded with waves of support for UBI and related policies (Widerquist forthcoming 2). These enormous undertakings require a great deal of political support. Social science experiments are usually too big to be funded by an everyday grant from a science foundation. The 1970s experiments were commissioned, not by private or public 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. Experiments in Namibia, India, Kenya, and two in the United States are all led or funded by private organizations with a strong interest in the UBI debate, although a mix of private and public institutional funding has been involved in some cases (Widerquist forthcoming 1).

Whether researchers like it or not, people on all sides of the UBI discussion all over the world will look to UBI experiments for information about UBI and sometimes for ammunition to use in debate. The experiments will affect the public discussion of UBI. People will seize on findings and say it implies X about whether UBI works or whether we should introduce it. The data will be used this way. The question is whether it will be understood and used appropriately; or misunderstood and abused.

To achieve the goal of enlightening discussion, people commissioning and conducting experiments need to know the local discussion well, but they also need to avoid overconfidence in their belief about how well they know it. Having read a few articles does not make you an expert. Journalists and opinion writers who have platforms to write about UBI are not necessarily experts on the UBI discussion, nor does most of the discussion go on in the pages of major media outlets. People commissioning and conducting experiments should not be tempted to believe that no one in the local discussion is interested in the

big questions that have not been explicitly stressed by prominent writers and speakers involved in the discussion. Ignoring the obvious and rational desire for anyone considering a public policy question to have answers to the big questions about it creates an opportunity for a demagogue to use that lack of information to spin the experiment's findings to their advantage.

To help bridge the communication gaps, the book by Widerquist (forthcoming 1) has to focus extensively on how limited UBI experiments are in answering the big questions about UBI. It also discusses the many communications barriers that make it difficult for researches to present results in a way that successfully raises the level of understanding of evidence among people involved in the public discussion of UBI. Therefore, the book has a lot of negative things to say that might cause some UBI-supporters to reject experiments altogether. This is not my message; the message instead is how best to conduct a UBI experiment and communicate its results once the decision to conduct an experiment is made. Experiments are happening; it's important to make the best of them.

Widerquist (forthcoming 1) also makes many specific recommendations, including strategies for conducting an effective test and for combatting spin and misunderstanding. Perhaps the best way to sum up my perspective is the following recommendation. Treat experiment(s) as a small part of the effort to answer the questions necessary to evaluated UBI as a policy proposal and to explain which unknown factors remain. This recommendation does not mean that experiments must be conducted in conjunction with many other research efforts to answer all these questions. It means that experiments in isolation cannot be interpreted as saying very much at all about UBI as a policy. The true value of an experiment is making a small contribution to this larger effort. For non-specialists to understand this: additional evidence has to be discussed, and the limits of experimental methods (and the overall effort to research a policy prior to implementation) have to be stressed.

In addition to many more specific suggestions, the book stresses four broad strategies for achieving this goal:

1. Work back and forth from the public discussion to the experiment. Anyone commissioning, conducting, or writing about experiments should respect the national or regional discussion of UBI. Find out what they can about what people most want to know. Design a study to oriented as much as possible towards the questions that are important to the local discussion, paying careful attention to the extent to which experiment can and cannot contribute to our understanding of those issues. All reports about experimental findings should relate the information to the big questions that are important to the local discussion. This strategy involves bringing in non-experimental data and calling attention to all experimental limitations, but it is necessary to help people appreciate the contribution an experiment can make.
2. Focus on the effects rather than the side-effects of UBI. Research projects have a way of focusing attention on the things they can measure at the expense of more difficult questions that might be more important to the policy issue at hand. For example, although the costs of UBI are important and more easily quantifiable, the most important question about UBI is whether it has the many positive effects on people's wellbeing that its supporters claim.
3. Focus on the bottom line. Although the public discussion varies enormously over time and place, the desire for an answer to the big questions is ubiquitous, and so I suggest focusing on what I call the bottom line: an overall evaluation of UBI as a long-term, national policy.[2] Experiments alone cannot provide enough evidence to answer a bottom-line question, but researchers can relate all of their findings to it. Virtually all UBI research has some relevance to that bottom-line evaluation, but citizens and policymakers, often need a great deal of help to understand those implications meaningfully, and even the best journalists are not always able to provide that help.
4. Address the ethical controversy. Researchers cannot resolve the controversy over the moral evaluation of UBI, nor should they try. But they do the public a disserves by ignoring it. They will do better to recognize the controversy and to explain what the findings mean to people who hold different ethical positions that are common in the discussion locally, and perhaps internationally too.

I wish I could say that this strategy will resolve this issue, but no effort to improve experimentation and communication will be perfect. A social science experiment is a very limited tool, and its implications are inherently difficult to understand. The bid to treat experiments as a small and incomplete part of a wider effort to answer all the important empirical issues about UBI will not even eliminate the need for caveats, although it will change the nature of the caveats involved.

There will always be gaps in understanding between the people involved in the discussion of such a complex issue and such complex evidence. If a non-specialist learns everything a specialist knows, they become a specialist. But experimentation and communication can always be improved and I hope

[2] UBI can, of course, be a regional policy.

that this research project makes a small contribution to that effort.

REFERENCES

Condliffe, J. (2016), "In 2017, We Will Find out If a Basic Income Makes Sense", *MIT Technology Review*, 19 December.

Deaton, A. and N. Cartwright (2016), *Understanding and Misunderstanding Randomized Controlled Trials,* NBER Working Paper 22595.

Teele, D.L. (ed., 2014), *Field Experiments and Their Critics: Essays on the Uses and Abuses of Experimentation in the Social Sciences,* New Haven, CT: Yale University Press.

Widerquist, K. (2005), "A Failure to Communicate: What (If Anything) Can We Learn from the Negative Income Tax Experiments?", *Journal of Socio-Economics* 34, 49–81.

Widerquist, K. (forthcoming 1), *A Critical Analysis of Basic Income Experiments for Researchers, Policymakers, and Citizens: The Devil's in the Caveats*, New York: Palgrave Macmillan.

Widerquist, K. (forthcoming 2), "Three Waves of Basic Income Support", in: Torry, M. (ed.), *Palgrave International Handbook of Basic Income*, New York: Palgrave-Macmillan.