

1

Designing Research

I Have a Question

HOW DOES THE WORLD WORK? That’s the big question, isn’t it? And it’s one for which our answers will always be ever so slightly incomplete.

That incompleteness is a curse, but also a blessing. Sure, we’ll never know everything perfectly.¹ But that also means that we’ll always have questions that need answering. For a certain kind of person, at least, answering those questions sounds like a pretty good way to spend your time on Earth. Maybe you’re that kind of person. I certainly am.

¹ Quantum mechanics is perhaps the most precise scientific field in existence, with accurate measurements and predictions out to more than a dozen decimal places. But even then, “accurate to fourteen decimal places” isn’t the same thing as “accurate.”

A RESEARCH QUESTION is a question that you have that you plan to answer, or at least try to answer, by doing research. Simple as that. Or, rather, as difficult as that. A good research question is well-defined, answerable, and understandable—those can be hard to figure out! We’ll talk more about this in Chapter 2.

For one example, let’s say that our research question is “does adding an additional highway lane reduce traffic?”

That’s a question about how the world works. Traffic is, unfortunately, a part of the world. And it’s something we could likely figure out with some research!

What kind of research? Well-designed research is research capable of answering the question it’s trying to answer. That seems simple but it actually requires quite a lot of thought and effort.

And that’s the real trick.

How can you do research in such a way that, when you’re done, you have an answer to your research question?

That’s what this book is about.

Empirical Research

“RESEARCH CAPABLE OF ANSWERING THE QUESTION IT’S TRYING TO ANSWER” could mean a lot of things, of course.

There are many kinds of research. You could look in books to see what people have already had to say about your question (“what do the traffic experts say about the effects of an additional highway lane?”). You could philosophically reason your way around the question (“if I assume people try to minimize their commute times, how would I expect them to respond to an additional highway lane?”). These are all forms of research.

This book will focus on *empirical* research, and specifically *quantitative empirical research*.

Empirical research is any research that uses structured observations from the real world to attempt to answer questions. So instead of trying to *reason* our way through what drivers would do if given an additional highway lane, we try to *observe* the choices that drivers take. Perhaps we interview drivers about how they make decisions. Or maybe we get a big data set of traffic violations, or of traffic flow numbers on highways.

Quantitative empirical research is just empirical research that uses quantitative measurements (numbers, usually). More data sets, fewer interviews.

QUANTITATIVE EMPIRICAL RESEARCH, like any kind of research, can be tricky! Measurements are difficult to take precisely. Statistics is a difficult field.

One particularly sticky problem with quantitative empirical research is that the numbers that we observe often don’t tell us exactly what we want to know.

After all, we probably aren’t interested in how much traffic there is on three-lane highways and on two-lane highways. We’re *probably* interested in whether we can make traffic go down by turning a two-lane highway into a three lane highway! But as much as we want them to, the numbers we have don’t actually tell us that right away. All we have are two-lane highways and three-lane highways. We don’t have a “what if” highway that tells us how much traffic there *would have been* if we’d made it one lane wider.

This problem constitutes a major headache for us researchers. If the numbers we have don’t actually answer the research question we have, what can we do?

Well, it turns out that, if you do it right, you often *can* figure out how to collect the right numbers, or do the right things to those num-

Empirical research. Research that uses observations from the real world in a structured way.

bers, to get an actual answer to our question. But it doesn't come free. We have to carefully design the right kind of analysis that will answer our question.

Why Research Needs a Design

WHY IS IT SO IMPORTANT for research to be properly designed? One way we can think about this is by looking at what happens when it's not.

Let's take our highways and traffic example. How might we go about researching an answer to this question? Our first pass might be to just compare traffic patterns on highways with more lanes against traffic patterns on highways with fewer lanes.

Seems reasonable. But then you do it and it turns out that more lanes seem to go along with more traffic! Surprising. However, why do those highways have more lanes in the first place? It might be that the busiest routes tend to be the ones that get expanded, and so it's no surprise that more lanes are associated with more traffic! Sure, maybe additional lanes *do* lead to more traffic.² But it takes research design to know that our first-pass analysis wasn't right, and figure out what to do instead.

A LACK OF SOLID RESEARCH DESIGN can be seen in the results, as well. Have you ever noticed, for example, how the studies you read about nutrition in the news can't seem to make up their mind? When I was a kid in the 90's, high-carb, low-fat food was what you were supposed to eat, and frozen yogurt and bagels both counted as pretty darn good for you. None of that is considered true today. And is a glass of wine a night good for you or not? Or coffee? Or butter vs. margarine? Or sugar vs. corn syrup?³ The underlying truth about what food is good to eat can't possibly be changing that much, but the scientific results sure do!

Some of this we can blame on the news hyping up studies beyond reason, or misinterpreting them completely. But some of this comes down to a lot of nutrition studies not having research designs that allow them to answer the question "what food will make you healthier?" Different studies seem to give different answers to "what food will make you healthier?" because they're not actually answering that question in the first place, even if they claim to! $2 + 2$ only has one answer, but if you're actually calculating something entirely different from $2 + 2$, you might well come back with an answer of 6, or 1, or -52. Then you wake up to a news headline reading that scientists have determined that $2 + 2 = -52$.

² And transportation researchers would generally say that more lanes at least leads to more driving! See for example Milam et al. (2017).

³ Ever noticed that some candy and soda brands advertise their use of "real sugar" as though not being corn syrup makes it a health food? This is only marginally related to what we're talking about here but boy do I hate it a lot.

Nutrition is a good field to pick on here because it isn't really the fault of the nutrition researchers themselves. Nutrition just happens to be a topic that makes good research design really elusive.⁴ So you end up with a field with shaky research design. And what does that give us? Inconsistent results that people have unfortunately learned not to pay all that much attention to today, because they know it might change tomorrow.

⁴ It's really hard to accurately measure what people eat, it's really hard to pick apart the effect of one food from all the other stuff people eat, it's really hard to separate out the effect of the food from the effect of the stuff that made you choose to eat the food, and so on and so on...

RESEARCH DESIGN IS HARD, and just because you want to answer a question doesn't mean there's necessarily a straightforward way of doing it. But the worst that could happen is that we'd figure out that the answer will be difficult to get. Then, at least, we'll know.

The best that could happen is that we *can* answer our question. And we do. And then we win a Nobel prize.

In This Book

THIS BOOK IS DESIGNED TO DO A FEW THINGS.

In the first half, it aims to teach you the principles of research design. Specifically, it will go into the ways that you can build an answerable research question, and then think about what kind of quantitative empirical research you could perform in order to answer that question. What would you need to measure? How could you be sure your method would actually answer your question?

Then, in the second half, it introduces you to some of the "toolbox" methods for causal research designs using observational data (i.e. answering a causal research question without running an experiment). These methods are very commonly used in modern research as they tend to be widely applicable in answering a wide range of research questions, and the assumptions they rely on are well-understood.

Hopefully, you can walk away from this book confident in your ability to craft a research project, figure out what kind of data you need to answer your research question, and figure out what calculations you need to perform on your data.