

## How to Formulate A Research Question:

A Guide for the Casual Researcher by Robert Fannion

So you're starting to wonder what to do about that Junior tutorial paper. The end of the semester seems far enough away, but 30 pages of research and argument suddenly seem like they might demand something more like five years and a government grant. What is a research project? How do you decide what topic to write on? Once you've decided on a topic, what kind of alchemy is involved in turning a topic into a question and a question into an argument? Well, help is on the way. You can write a paper that makes an interesting and original contribution to the social sciences, and it's easier than you think. Just keep your eye on the ultimate goal of research work and follow these easy steps and you'll have a roadmap to a good project in no time. Oh, you'll still encounter frustration and panic and nights where you wake up in a cold sweat and wonder if what you're doing makes any sense, but that's all part of the fun.

Start with a puzzle; a good puzzle is almost always the foundation of a good research project. For the social sciences, this usually means starting with something that you observe in the real world and want to be able to explain. Specifically, think about the issues that you care about as the starting point of where you should look for phenomena to explain. If you care about development in Africa, then look either at events in Africa or at things elsewhere in time or space that you think might give you insights into what you care about today in Africa (e.g. the growth of cities in Europe in the 19<sup>th</sup> century presented many of the same problems of providing clean water, sewage, and transportation to urban residents that are being faced now in Africa – examining another time and place are often very useful in helping you to develop arguments about current events). Look around at how people or organizations operate through the eyes of a social scientist and you will see lots of interesting patterns. *These patterns of behavior and the ways in which they interact are the subjects that we often want to explain.*

But not everything that you observe in the social world is necessarily a puzzle – and certainly not an interesting one. It is sometimes useful to think about puzzles as coming in two flavors: empirical puzzles and theoretical puzzles. This is a bit of a simplification, actually, because all puzzles are both theoretical and empirical at once, but for how we approach them this classification can be useful. Let's start with theoretical puzzles. What makes a social phenomenon into this kind of a puzzle? The answer is simple: puzzles exist when it isn't obvious *why* something works the way that it does. This is where social theory jumps out of the dusty pages of journals and becomes what it is meant to be: an intellectual tool that we use to take apart and interpret reality. Social scientists have lots of theoretical tools that tell us how the world "should" work, and when the world seems to actually operate more or less in accord with our models, then it gives us a warm feeling of satisfaction. But, because the theories that we use are always both *contingent* and *incomplete*, there are lots of situations in the real world where people and organizations don't behave in the ways that social science would expect. When this happens, there is something to explain, and you have a puzzle.

Theoretical puzzles are especially interesting to social scientists because they challenge our existing tools of explanation and force us to either modify them or develop new ones. So let's take an example. Imagine that you'd just read Gerschenkron and decided that you wanted to examine economic development in the United States. Since

the U.S. industrialized more or less in the late 19<sup>th</sup> century, Gerschenkron might lead you to expect that the capital requirements associated with the technologies of the time would have produced a system dominated by large universal banks like those in Germany. For some reason, this didn't seem to happen. Is Gerschenkron's theory simply wrong? Is it incomplete? Is there something unusual about the United States that would interfere with the causal process outlined by Gerschenkron?

Here we have a disjuncture between on the one hand what a theoretical perspective would lead us to expect and on the other what actually occurred. That's a theoretical puzzle – the world didn't work the way that Gerschenkron says it should have (at least the way that a very simple take on Gerschenkron suggests it should). But we can also invert our perspective and treat the case of American banking as an empirical puzzle: how did America finance its industrial development? What determined the shape of the institutional structures that were developed for this purpose?

A paper that starts from Gerschenkron and asks why the United States didn't develop large universal banks might make a big or a small theoretical contribution. It could conclude that there were some unique circumstances in America that prevented universal banks from developing and instead diverted investment into another set of institutions that accomplish the same function. Alternately, it might use the American case to offer a more fundamental critique of Gerschenkron's argument by showing that the factors that determined financial intermediation and capital concentration in the US were completely unrelated to the requirements of dominant production technologies. Even in the relatively unambitious former case, you are adding a new dimension or a boundary condition to Gerschenkron's argument ("his argument does not seem to apply when the following conditions are present..."). This helps to improve his argument and ultimately refines and improves the theoretical toolbox available to social scientists when we try to figure out the world. From a certain perspective, that's what we want you to accomplish in your research papers. And it's actually pretty cool.

Discovering why a model fails to explain something and creating a model that would better fit your cases is the key to turning a puzzle into an argument. Theoretical puzzles can challenge how theories are *applied* – adding epicycles to help account for interesting *variations* in a case that you have studied – or they can challenge the basic *behavioral or institutional assumptions* on which existing theories are based. That last kind of puzzle can be the most intellectually exciting, but even puzzles that suggest that we need to add footnotes or caveats to existing theories can provide the basis for worthwhile research.

Now let's turn more explicitly to empirical puzzles. Empirical puzzles are simpler intellectually, but they can be more practically rewarding than primarily theoretical ones. An empirical puzzle is usually based on something that you find significant in the real world that can be explained using some combination of existing theories. Why is corruption so common in African governments? Why is the consumer protection movement so weak in Japan? Why doesn't the United States government guarantee universal basic health for its citizens like other industrialized countries? Why do some firms choose to build factories in economically underdeveloped countries while others continue to produce goods in the developed world? In these cases, you aren't starting with a clear disjuncture between one specific theory and reality (though if you look closely at your assumptions in posing questions like the above you'll usually end up finding a theoretical puzzle hidden inside an empirical one). Instead, you are asking what

set of theories in the social sciences can best explain a phenomenon that we find interesting because it affects things that you (and the audience for your project) care about.

All right, so say you have found a puzzle that you find interesting. The next step is to put that puzzle in a theoretical context. Social phenomena are immensely complex, so a big part of our job as researchers is to simplify them enough to make generalizable arguments. To help us walk through the process, let's take an example. Say that I wanted to explain why women entered the labor force in the United States from 1950-1970. Obviously, there are as many specific, human answers to this question as there were women who started cashing paychecks during this era, but that immensely complex truth doesn't help me. So I begin by thinking about how I can simplify this reality enough to even be able to grapple with it, and to do this I turn to social theory. For our hypothetical puzzle, I might start by thinking about the problem in economic terms. Let's assume that women are rational, utility maximizing individuals who make the decision to seek paid employment based on their weighing of the monetary return to their labor relative to what they were presumably doing before – possibly non-paid labor in the home. If I started with this set of simplifying assumptions, glossing over the immense diversity of why women entered the labor force, I would look for forces that raised women's wages during this period or made non-paid labor in the home relatively less rewarding. I might look at unemployment levels and wage data and find that this period saw very high demand for labor in a rapidly growing economy, leading to greater opportunities for women.

Alternately, I might start with a set of assumptions that are more sociological than economic. I might think about the independence and social rewards of earning money through paid labor and argue that women have always wanted to participate in the labor force on equal terms with men, but that laws and customs had prevented them from doing so. Based on this set of assumptions, I might look not at changes in wage levels, but at employment or anti-discrimination laws, evidence of changing social attitudes, or changing patterns of family and marriage. The social constraints on what women were *allowed* to do would be the center of my model in this case, with less emphasis on their rational choices but more on what *constrained* those choices.

Let's step back and think about what we are doing when we think about a puzzle in terms of simplifications like this. In order to frame my research question, I have taken my puzzle and started to search through the social sciences for the theoretical tools that I think are most useful in understanding it. Using these tools, I construct a simplified model of the *causal process* that I want to understand. In the example above, I would read work on either the microeconomics of labor markets or on the social effects of discriminatory attitudes and the practical impact of labor law. This exercise in translating your puzzle into a model will appear in your written project in the form of a literature review or theoretical section. The goal in this section of your research project will be to explain the way in which you have chosen to model social reality and then to justify that model as being superior to the alternatives. To continue the example above, my theoretical section would explain why I chose to examine labor force participation in terms of legal and cultural discrimination *instead* of modeling it as the aggregation of individual women's rational choices about the trade-offs between paid labor and unpaid work in traditional social roles as wives, mothers, etc. Explaining to your reader *why* the

theoretical framing you have chosen is the most appropriate involves taking on possible criticisms and alternate viewpoints and presenting evidence that the simplifying assumptions *you* have chosen to make are more realistic or better capture social reality than the frameworks that you reject.

Once you have chosen how to frame your puzzle in theoretical terms, you are ready to make some specific causal claims about your topic. Deciding what claims to make is often difficult, and you are very likely to come back to this section of your project once you have done empirical research and change those claims. Don't worry about this! At an early stage in your project, you shouldn't be afraid to make an argument that is fairly bold and then come back and modify or qualify that claim later. As you do this, you will also want to perform some preliminary empirical work to survey the facts about your topic in order to make your theoretical claims plausible. This won't require too much effort, and in many cases can be done by examining secondary literature about your topic, but it is important in making sure that you are generating theories that aren't completely at odds with the reality you want to understand. For social scientists, this kind of a constant reality check is important, because it is so easy to become lost in theoretical debates that you lose sight of the real world – as Thoreau said, *you are in danger of becoming a tool of your tools, that is, a servant of your simplifying assumptions instead of their master.*

In our hypothetical research project, I might look at some secondary literature on the postwar American economy and think about the importance of cultural attitudes toward working women and decide to argue that women entered the labor force in the postwar era because of an explosion in demand for labor in the kinds of jobs that women had traditionally performed. The growth of the consumer economy after World War II created huge numbers of jobs in *retail services* and in low-paid *clerical and office positions*, things that we might call “pink-collar” jobs. Women had been office workers and retail clerks since the late 19<sup>th</sup> century, and like nursing or education these jobs were seen traditionally as “women’s work” (often expected to be a temporary position held until a woman was married and exited the labor force). I might argue that the postwar American economy generated huge numbers of what were traditionally seen as female jobs, and that this gave woman an opportunity to enter the labor force without facing the degree of cultural backlash that would have emerged from women taking blue collar or white collar jobs traditionally seen as the preserve of men.

Let's review what we have done intellectually so far. We start by identifying something interesting in the social world, either something that seems to contradict the way our theories tell us the world is supposed to work or something that is important enough in itself to deserve explanation. Once we have done this, our next step is to make some basic decisions about how to model this social phenomenon. What kind of simplifying assumptions can we make without losing sight of the forces driving what we want to study? What existing literature addresses these kinds of assumptions and this kind of causal mechanism? How close is our argument to the branch of existing social theory that you have chosen? After putting together a general theoretical frame for your topic, the next step is to develop a specific causal argument based on this framework that can be used to “solve” your puzzle. If your puzzle asks why, this is the stage at which you make a first cut at saying “because.”

The next step in the process will guide your empirical work – the data collection and analysis that we generally think of as the core of research in the social sciences. Think carefully about your explanation and ask yourself what the empirical implications of that explanation should be. Put simply, if your causal story is correct, what facts should you be able to find that support it? We call those the “**observable implications**” of a theory, and these observable implications tell us what data we need in order to support our argument. At this stage, it might be helpful to make a list of the effects your causal story would have on the real world. To continue our ongoing example, I might decide that if my causal explanation about women entering the workforce is correct, I should be able to find evidence that “pink-collar” segments of the economy were expanding rapidly and that the numbers of jobs being generated in retail and clerical work were in fact being taken by women. I might also decide that it is important to gather evidence about social attitudes toward women in different kinds of jobs; maybe I could survey magazines aimed at women in the 1950’s and 1960’s to find out whether the articles deal more with issues that affect working women or discuss certain kinds of jobs as one step in finding out how attitudes changed during this period.

The ideal behind this is to generate testable hypotheses that will support your causal argument. These hypotheses generally take the form: if X is true, we should observe Y. You will probably want to come up with a list of these hypotheses that cover different aspects of your causal story, even though in your final project you will end up focusing on only a few of them.

With this done, you are ready to go out and look for evidence. Your research at this stage will consist of finding the evidence that fits in your hypotheses and allows you to support the causal argument you have made. This might be statistical evidence, observations that show that X and Y are correlated in the way that your causal argument suggests that they should be. It might be evidence based on interviews or survey results. It might be inference based on how your causal process affects other members of society. We are all prisoners in the social sciences of our ability to observe the real world, so we tend to make the kinds of arguments that can be supported with the evidence that we have at hand. In addition, we are often lazy and would prefer to make an argument that can be supported by data that has already been collected by someone else. When you are thinking about how to generate hypotheses with observable implications, this question of data availability should be present in your mind.

In our example above, I might decide that I need to find evidence that popular attitudes toward women working actually varied by the kind of job they held – maybe popular magazines or government agencies collected some survey data on attitudes toward women in different professions during this period. Perhaps I could interview women who actually entered the labor force in those years and ask questions about what opportunities were available to them, what jobs they even thought about applying for and why. I might also want to look at aggregate level data from the Bureau of Labor Statistics. Is it true that large numbers of clerical and retail jobs were created in this era? Is it true that most of those jobs were held by women? Were most of those women married? How old were they? I might also look at industry journals for sectors of the economy where these jobs were being created. Perhaps the American Banker’s Association published articles on hiring women as bank tellers. Perhaps large department

stores published something on finding salesladies for new stores being opened in new suburbs.

In making these decisions about what kinds of evidence to seek out, one valuable resource is other work on similar topics. What kinds of evidence have other people making similar arguments used? Looking at journal articles and books can give you an idea of how other people have solved exactly the problem that you are facing: what kinds of evidence can be used to test hypotheses about your topic. Looking at these articles can also help you by pointing you to specific data sources that you might not have known existed. Perhaps there is an oral history project at the Smithsonian that interviewed working women in this era. Perhaps an international organization like the ILO compiled data on American women in the labor force in the 1950's to determine whether they were joining unions. By looking at what other researchers have used to support their arguments, you can save yourself the reinvention of the wheel in actually collecting and analyzing data.

Overall, the goal of this part of the project is to put together a case that supports your causal argument and demonstrates that this argument represents a better explanation than the most obvious alternatives. In our ongoing example, I might decide that in order to counter the microeconomic argument, it would be useful to find wage data that show whether women were being paid more in this period in order to evaluate the claim that higher wages pulled women into the labor force.

This should set you on course for developing an argument and finding support or it. Later in the semester, we will talk more about how this comes together into a paper or a thesis.