How to Think About Your Research Paper

Kieran Healy∗

February 14th 1999

Most of the readings for this course are empirical studies of some aspect of economic life. Their authors began with an idea or theory they wanted to examine, and they designed a study that would let them field-test this idea in the real world. The paper requirement for this course asks that you do much the same sort of thing yourself, on a smaller scale. This memo is meant to briefly clarify some of the principles that guide the empirical research that sociologists do, and to discuss some of the methods they use to do it. You’ll probably use one of these methods to collect the data for your paper. First we’ll look at some of the basic concepts that guide empirical research. Then we’ll discuss four methods — surveys, interviews, participant observation and documentary analysis.

What does a research paper look like? The best ones ask a well-defined and interesting research question; they present competing hypotheses drawn from more general theories; and they test these hypotheses against good empirical data that have been collected using some systematic method.

∗Email at kjhealy@u.arizona.edu. Originally written (by me) as a handout for Soc 212, Social Relations in the Economy, taught at Princeton by Viviana Zelizer. I have rewritten it somewhat to help it stand on its own.

1This memo draws in part on some ideas presented in Designing Social Inquiry: scientific inference in qualitative research by Gary King et al. (Princeton University Press 1994). This is an informative and accessible book that you might find useful to read in more detail. Chapters 1, 4 and 5 are especially worth reading. I’d also strongly recommend you take a look at three other books: Arthur Stinchcombe, Constructing Social Theories (Chicago 1987), Wayne C. Booth et al., The Craft of Research (Chicago 1995), and Howard Becker’s Tricks of the Trade: how to think about your research while you’re doing it (Chicago 1998). Throughout the memo, references to works mentioned in each section will be found in a footnote like this one at the end of that section. Papers and books discussed in the text are taken from the syllabus for the course.
A research question

A research paper should answer an empirical question, that is a question about something that happens in the world. How do people find jobs? How do auctions really work? Why don’t Welfare-reliant mothers go to work? For sociologists, research questions are usually asked about some aspect of the world that we don’t know about, or haven’t investigated closely before. Christine Williams’s book *Still a Man’s World* is an example of this kind of research. We also ask a research question about something we *think* we know about, but go on to give a surprising answer. Bruce Western and Katherine Beckett’s study of how the penal system affects the labor market is a good example here. Or our choice of question may be guided directly by theory: we deliberately choose to study something that will help us confirm, modify or falsify a theory that already exists. Daniel Kahneman *et al.*’s experimental test of the assumptions of rational choice theory shows one way to do this sort of research. A good research question is well-defined, clearly stated, and addresses an empirical or theoretical issue.

Bear in mind that the rather dry language of formal social research — competing theories, formal hypotheses, and so on — can make it easy to forget the point of the enterprise. We want to ask and answer important and interesting questions. But the big, sweeping questions that motivate whole fields are impossible to tackle properly in a short paper. It is possible to write interesting papers that don’t over-reach, and to combine ambition to address big questions with the discipline to focus on specific aspects of them. This issue is well discussed by Murray Davis. His paper is titled “That’s Interesting!” which is exactly the response you want to get from people when they read your paper.²

Theories and hypotheses

Theories explain things. If a theory is to be useful, it has to have some observable implications about the world. It should tell us what will happen

in the specific instances that make up whatever general case it claims to explain. Are weak network ties good for finding all jobs, or just some jobs? Do people make rational choices in all areas of their lives, or just some of them? If only some, then which ones exactly? The theory should tell us these things — it should be clear about what it explains. Even if a quite general theory doesn’t have anything to say about some specific case, it should still be clear enough for the researcher to work out what it implies about that case.

These implications help us distinguish between relevant and irrelevant data when we come to test the theory. They should also allow us to discover instances where two theories predict opposite (or at least different) things. Such cases are especially useful in empirical research, because at least one of our theories has to be wrong (about that particular situation). Good research should try to find empirical cases where different theories make particularly strong predictions about what should happen. That is, researchers try to find decisive tests of the available theories.

There are different ways to do this. For example, Charles Smith’s work on auctions suggests that people do not behave as rational choice theory says they should in these situations. Auctions are an interesting site for research because, as Smith points out, in many ways they are events where rational decision-making should occur exactly according to the theory. By contrast, Richard Posner believes rational choice theory to be broadly correct. His “Sex and Rationality” sets a strong test for the theory in a different way, by applying it to an area (sex, love and marriage) where most people intuitively believe it could not possibly be true. If he can show his theory can account for what goes on in this area, this is strong evidence for the robustness of the theory.

Many of the articles and extracts in the reading packet take a general theory and work out what it predicts about some specific bit of social life — that is, they formulate clear, comparative, falsifiable hypotheses from the theory. Where they try to test more than one theory, they try to ensure that the hypotheses make different predictions about the same thing — that is, they compete with one another. In this way, the data can test the hypotheses and kill (or confirm) two theoretical birds with one empirical stone.3

Empirical data, and its relationship to theory

Empirical data is just well-ordered information about the world that has been collected using some systematic method. The trick is making it well-ordered. People sometimes take “empirical” data to mean “statistical” data. This is a mistake: “empirical research” just means the study of things that happen in the world. It can be done using a wide variety of data and methods, from official statistics to personal interviews. You should also be careful not to think that statistics are more objective than, say, interview data, just because they are numbers. Although different kinds of data are suited to different sorts of problems, the most important thing about any dataset is the way it was collected. A carefully designed, systematic set of interviews with a proper sample of the population will tell us much more about the world than a statistic calculated from a biased source or botched survey.

We said earlier that a good theory should have observable implications about the world, and that we should be able to state these implications as hypotheses. That is, the theory should make a clear claim about what we will observe under specific circumstances. A hypothesis is usually a claim about causes and effects: Variable X causes effect Y (under certain conditions). In *The Gift Relationship*, for example, Richard Titmuss tries to show that a market for human blood causes the supply to become contaminated, because blood markets attract suppliers who are (for various reasons) more likely to be infected with a blood-borne disease. In “Conflict Management, Honor and Organizational Change,” Calvin Morrill argues that changes in the management structure and business environment of a large toy company led to the emergence of a new (and rather odd) “culture of honor” within the organization. In “The Penal System and the Labor Market,” Bruce Western and Katherine Beckett argue that sentencing and imprisonment policies have large (and under-appreciated) effects on the American labor market.

Though they use different methods, each of these papers claims a link between a cause and an effect, and then tries to show how the two are related. To do this, they must (i) show that the alleged cause could, in fact, bring about the observed effect; and (ii) as far as possible, eliminate (“control for”) the possibility that something else might be causing the effect they are interested in. Meeting these standards can be difficult — even the best research very rarely settles causal questions like these once and for all. If you’ve ever taken a course in the philosophy of science, you’ll know that inferring some real, underlying cause from a set of observations is a deep problem. The same goes for deciding what counts as an explanation. We’re
not about to solve these problems here. Social science data is almost always “noisy” and could usually do with more controls. However, there are some conventional standards that we can aspire to. A careful design is what makes a convincing research paper out of an interesting question.\(^4\)

**Variables and variance**

The classic means of isolating a cause is the controlled, randomized experiment. We don’t use this method very often, simply because you usually can’t do experiments to resolve the questions social scientists are interested in. In an experiment, the principles guiding social research are very clear, so it’s worth briefly discussing the experimental method before we see how these principles carry over into other methods.

An experimental research design isolates one potential cause and one observed effect in an elegant way. Say you wanted to see whether eating chocolate made people forget their names. To test this idea experimentally, pick 100 people, write down their names, and then randomly divide them into two groups. Give Group A some chocolate and make sure they eat it (this is the experimental group). Don’t do anything to Group B (this is the control group). Then ask everyone in both groups to tell you their name.

If no-one in the chocolate-eating group knows their name, but everyone in the control group does, then this is good evidence that eating chocolate makes people forget their names.\(^5\) The evidence is good because the chocolate is the only systematic difference between the experimental and the control groups. It might be that many variables — race, gender, height, eye color, political affiliation — affect the ability to remember one’s name. But because you randomly assigned people to the two groups, these characteristics won’t have any systematic effect.\(^6\) Any effect you do observe is therefore attributable to the one thing that does vary between the groups.


\(^5\)How many of the 50 people in the experimental group would have to forget their name before you were convinced that the chocolate had an effect? All 50? Only 49? How about 35? We have statistical tests that help us quantify whether a difference of this kind is due to random chance or a genuine underlying cause — this is what the phrase “statistically significant” means.

\(^6\)That is, the random assignment of subjects to the different experimental groups controls for the effects of all the variables except chocolate consumption. Hence the name “controlled randomized experiment.”
namely chocolate consumption. Notice that it’s this variation between the two groups that allows us to infer the cause.

Unfortunately, the social world is not a controllable experiment. But the same principles of explanation apply: control for the variables you’re not interested in; make sure what you are interested in varies. If you look at the sample paper topics, you’ll see that most of them suggest that you carry out some comparison between two otherwise similar things that vary in some important respect:

“Choose two products, media, or organizations that vary significantly in how much they are oriented to a particular local market.”

“How many magazines (articles, letters to the editor) with etiquette books, and websites for teens.”

It should be clearer at this point why you’re being asked to make comparisons like this. **Comparison implies variation** on some dimension, and **variation gives us explanatory leverage**. Look again at Granovetter’s *Getting a Job*. He was forced to restrict the variation on some variables to zero, but he tried as hard as he could to get others (the ones he was interested in) to vary as much as possible. He deliberately excluded women from his survey. He only looked at Professional, Technical and Managerial workers. Almost everyone in his survey was white. These restrictions allowed him to see more clearly the effect of the causal variable he was interested in — the strength of people’s network ties.7

To make a causal claim, there needs to be variation both on the **independent** variables (i.e., the things you think are causes) and the **dependent** variable (the thing you think is an effect). If you gave a chocolate bar to

---

7 Note that this way of controlling for variables means that, strictly speaking, Granovetter’s conclusions only apply to white, male, professional workers. He was forced to restrict his sample in this way for logistical reasons. Ideally, Granovetter would have interviewed many hundreds of people — men and women, professional and manual, and so on — but he didn’t have the time, the staff, or the money. Having many people in each category would have approximated experimental randomization, and made his conclusions apply to a much wider population. Unfortunately, conducting that many interviews wasn’t possible, so he was forced to settle for the second best option of deliberately eliminating groups from the sample. This method is defensible theoretically as well as practically. As long as you have strong reason to believe that, so far as your study is concerned, those you exclude are no different from those you include, it shouldn’t be a problem. Granovetter’s general conclusions tacitly assume that women’s networks work the same way as men’s. Perhaps they don’t; in which case his findings raise a new research question about women.
everyone in the chocolate experiment (i.e. to both the experimental and the control groups) you would be criticized for having “no variance on your independent variable.” If everyone gets a bar, you can’t say if it’s causing anything, because you have nothing to compare it to. The same goes for the dependent variable. If everyone in a survey says they vote Democrat, then it’s impossible to make a causal argument about what independent variable causes people to vote democrat. Intuitively, we can infer causation only if two variables are associated with one another, meaning that when one changes the other does, too. “Lack of variance” just means that something isn’t varying, or changing — hence, no causal inferences are possible.

People often spot problems with lack of variance on the independent variables (causes), but a common mistake made with dependent variables (outcomes) is a little more subtle. Say I wanted to find out what makes people commit crime. It might seem obvious to go to the local state prison and interview some criminals. I carefully conduct 100 interviews and find that 70 per cent of my interviewees had a relative who supports the Republican party. I therefore conclude that having a Republican in the family causes one to commit crimes.

Of course, this is an incorrect finding (probably). If I sampled the entire population of the US, I’d likely find that 70 per cent of Americans had a Republican somewhere in their extended family, too. What’s happened is that the way I selected my sample is systematically related to the outcome I’m interested in. I was interested in criminals, so I went to the local prison. But by only interviewing prisoners, I corrupted my sample: there’s no variance on the outcome, so I’m likely to make all sorts of false inferences. I can’t tell whether my findings about prisoners are different from what’s true of non-prisoners, because there aren’t any non-prisoners in my sample. This error is known as sampling on the dependent variable. The problem might seem obvious from the example, but it can be a subtle error. Often, our sample selection method or data collection method can be systematically related to the dependent variable in ways that we don’t know about.

Choose an appropriate method

A different, but relevant, virtue of Granovetter’s study is that the method he chose to collect his data was appropriate to the question he wanted to ask. He was interested in how people found jobs, so he put together a postal survey and conducted about a hundred personal interviews. This might seem obvious, but the point is that different methods are good for different things. A corollary is that you should think about whether the data you plan to collect can really speak to your theory at all. For example, Morrill
did intensive ethnographic work to collect his data on the corporate culture of a toy firm. The strange vocabulary and actions of the managers would not have shown up in a documentary analysis of company records, or even in a detailed questionnaire about corporate culture — he wouldn’t have known what to ask about. Similarly, Western and Beckett would have been unlikely to have reached their conclusions about the labor market had they gone and observed a prison for a year.

For the rest of this memo, we’ll look briefly at some of the methods you might use to collect your own data, focusing on their strengths and weaknesses.

**Sample Surveys**

We often want to find about things about some particular population of individuals — welfare mothers, job seekers, citizens of the United States — but we don’t have the time or the money to ask every one of them for information. A survey draws a sample from a larger population and asks them the questions we’re interested in. If the sample is collected properly, then the results can be remarkably accurate. Newspaper polls about President Clinton’s popularity are collected by asking only about a thousand people what they think, but we know that they’re accurate about the views of 270 million Americans to within about three per cent either way — a pretty small margin of error, given that we only asked about 0.00003 per cent of the population.

This principle extends downwards: you can get interesting information about what students (as a whole) think by only sampling a few dozen students — as long as you do it carefully. A moment’s thought should make you see that all sorts of biases can creep into a sample. If Gallup took its thousand poll respondents from your freshman class, the results could not be validly extended to the entire U.S. population. The freshman class is a sample of the U.S. population, but it’s not a representative sample. When running a survey, you need to think about the population you’re actually selecting from.

Surveys are good for eliciting simple opinions, reports of recent behavior or past actions, and other well-defined and easily-recalled attitudes. Surveys often try to discover how attitudes or activities are associated with specific groups. In the general terms we just used, the independent variables are usually things like gender, ethnicity, age, religion, occupation and so on. Most of these can reliably be collected from a survey, too. The dependent variables are more often questions about political or social attitudes, beliefs,
and so on. As before, we want the same principles of variance and explanatory power to apply. From this perspective, the worst survey questions are the ones that everyone gives the same answer to. Well-designed survey research can tell us a lot about a large population, at the relatively small cost of 40 or 50 short interviews. Responses can be coded into numbers and analyzed using statistical methods.

There is a knack to asking questions that will elicit accurate responses. It’s easy to accidentally or deliberately structure a survey so the respondent is channeled towards giving a particular answer. For example, some of Microsoft’s ideas about survey research came out during their anti-trust trial. Microsoft corporation commissioned a survey from an independent research firm, but with this request from Bill Gates attached: “It would help me immensely to have a survey showing that 90 per cent of developers believe that putting the [Microsoft Internet Explorer] browser into the [Windows 98 operating system] makes sense.” The firm carrying out the survey went and asked a 350-word question that detailed the benefits of this integration, without mentioning that there might be any disadvantages. It’s therefore not surprising that 85 per cent of the respondents said they’d be in favor of it.

Surveys are not as useful for pinning down people’s views on questions that don’t have simple yes or no answers, or on opinions that are very sensitive to the way they are elicited. (Though of course this means they can be used to show which opinions are sensitive in this way.) Neither are they so good for second-hand reporting — you won’t get accurate answers to questions like how much housework a respondent’s mother did, for example. People don’t remember these sorts of things reliably. Even worse, people are often unreliable reporters of facts about themselves. People are not good estimators of things like number of hours spent studying or working, number of books read in the past year, or amount of TV watched in the last week. You should be wary about the validity of survey responses for these sorts of questions.

There are many other wrinkles to survey research, from ways of getting people to fill the thing out (mail surveys don’t get returned, no-one wants to talk to you on the street), to problems with interpreting the answers you get. To learn more about survey research and design, take a look at Earl Babbie’s *Survey Research Methods* (2nd ed, Wadsworth 1990), or Arlene Fink and Jacqueline B. Kosecoff’s *How to Conduct Surveys: A Step by Step Guide* (Sage 1998).
Interviews

Interviews are used as a research tool in one of two cases. First, when we want to find out about something in more detail than a questionnaire-based survey method can tell us. Perhaps the topic is sensitive or difficult to discuss — or just complicated. Second, when the people we are interested in form quite a small category, and they have specialized knowledge or experiences. Perhaps we are interested in how transplant surgeons view their work, or male librarians. In both these cases, interviews allow the researcher to ask more detailed questions, obtain more nuanced answers than surveys allow. We can also pursue lines of questioning or potentially interesting topics that emerge in the course of the interview itself. It may be that we don’t know what we’re looking for until we hear it. If so, then we won’t know the right question to ask in a survey.

Interviews usually need to be taped and transcribed, and this is a very time-consuming process. In the absence of subpoena powers, it’s much harder to get someone to submit to detailed questioning than it is to ask them a few quick questions outside the dining hall or on the street. Carrying out enough interviews to do any sort of statistical analysis is usually quite difficult. For the purposes of a short paper, an interview where you take detailed notes (or take notes from a recording of the interview) will suffice as your data.


Participant Observation

There’s more to society than values, attitudes and opinions. People also do things. Studying what they do and the way they do it demands a different sort of method. In participant observation (which in its more intensive, long-term form is also called *ethnography*), the researcher observes what’s going on by participating in it. While collecting the data for *Charismatic Capitalism*, Nicole Woolsey Biggart didn’t just interview people who sold for Mary Kay and Amway, she went to their meetings and attended their conventions. (I’m not sure if she sold any products, but she might well have done.) Charles Smith spent a lot of time attending auctions and observing
what people did, not just asking them about it afterwards. Although we don’t read his work in this course, the sociologist Erving Goffman was a brilliant observer of the subtleties of human behavior. Read his *The Presentation of Self in Everyday Life* (Anchor books) for a sense of how rich a method participant observation can be.

Participant observation and ethnography pose a number of challenges to the researcher, in particular how to balance participation against observation. Some of these issues are visible in Philippe Bourgois’s *In Search of Respect*. The strength of the method is its ability to give us a clear picture of some practice, or chunk of social life, that is either difficult to study at all or is routinely ignored as undeserving of consideration. Bourgois’s crack houses are an obvious example of the former kind; the paper topic that suggests spending time in ethnic shops and large chain-stores is an example of the latter. The differences in behavior and interactions that you would likely find between a local ethnic market and a large corporate outlet would be hard to get at through interviews or surveys.

Observing interactions in public settings generally does not pose any ethical difficulties. But even small-scale participant observation studies may questions about your research subjects’ rights to privacy. Be sure to check the Human Subjects Committee regulations if you plan to carry out any elaborate observational studies.

Note that the suggested paper topic incorporates the participant observation method into a comparative research design. It’s hard to overemphasize the importance of this distinction. Although different methods are used for different questions, the principles guiding research design are the same whatever the method. For practical advice on how to conduct participant-observer studies, see John Lofland and Lyn H. Lofland’s *Analyzing Social Settings: A Guide to Qualitative Observation and Analysis* (3rd edition, Wadsworth 1995).8

**Documentary analysis**

How do attitudes and behavior differ today from those held fifty or a hundred years ago? Obviously, we can’t travel back to the late nineteenth century to find out what people thought. Instead, we need to look at what they left behind. Advice books, magazine articles, etiquette manuals, govern-

---

ment reports, personal diaries, newspaper columns, and virtually any other written product of a culture can be used as a window on its attitudes, practices and preoccupations. By extension, the products of our own or any other contemporary societies can be examined in the same way. (That is, we can carry out cross-sectional comparisons — i.e., across space — as well as longitudinal ones — i.e., over time.)

Even where we are able to carry out opinion surveys or interviews, “cultural products” like magazines, TV programmes or books often provide evidence about the content or orientation of a culture, over and above the individual opinions of the people who make it up. The idea here is that there is more to a culture than the sum of people’s opinions. There’s a strong sense in which our culture exists outside of ourselves, and exerts an independent influence on what we think and do. Often, a set of cultural norms or expectations may cause people to act in a certain way, even though very few people really believe them. Documentary analysis can get at these dominant views or general expectations in a way that other methods cannot. For example, Kathy Peiss’s “Making Faces” looks at the rise of the cosmetics and beauty business in nineteenth century America. Note that her “comparative axis” is internal to that time: she looks at how cosmetics marketing differed along racial lines, and how the cosmetics industry developed standards of beauty that it tried to turn into cultural ideals. Similarly, one of the paper topics suggests that you use magazines and advice books aimed at teenage markets to study the relationship between monetary transactions and intimate relations.

A particular danger of documentary studies is that they tempt one to over-generalize. Drawing conclusions about a whole society on the basis of a few magazine articles or an episode of Friends is a dangerous business. On the other hand, both Peiss’s work and Viviana Zelizer’s The Social Meaning of Money show what you can get from this approach. As with any method, the power of the results will depend on to the degree to which you collect the data in a focused, systematic, well-designed way.

If this sounds a little odd to you, read Deborah Prentice and Dale Miller, “Pluralistic ignorance and alcohol use on campus: Some consequences of misperceiving the social norm,” Journal of Personality and Social Psychology 1993 Vol 64, number 2: 243-256. The research for this paper was carried out at Princeton University, incidentally.

References

The following is a list of the books about research methods mentioned in this memo, as well as a few others you might find useful while writing your paper. The library should have all of these books.


Becker, Howard S., *Tricks of the Trade: how to think about your research while you’re doing it* (Chicago 1998).


Ragin, Charles, and Howard S. Becker, *What Is a Case? Exploring the


