

“Asking Interesting Questions” Prepared for Sage Handbook of Research Methods in Political Science and International Relations. Robert J. Franzese, Jr. and Luigi Curini, Editors. *Forthcoming*.

By

William Roberts Clark¹
Texas A&M University

Good research is driven by impatience with bad answers to interesting questions. But where do interesting questions come from? Since this is the opening chapter of a handbook on research methods, it is imperative to point out at the start that there is no “method” to asking research questions in the sense of a cookbook that you can follow that will lead, inexorably, to scientific discovery. There *may* be a scientific method for evaluating answers, but there is certainly no scientific method for asking questions or generating answers. And there is certainly room for a lot of creativity in developing interesting and enlightening research designs, and serious shortcomings to “cookbook” approaches.² Karl Popper, (1962, 2003) for example, argued that science begins after a scientist has conjectured an answer to a question. The scientific method, therefore, is more (perhaps only) useful in evaluating answers to questions. Generating questions and answers, in contrast, is as much an art as it is a science.

But that is not to say that the process is random or lacks structure. Thomas Kuhn (1962) says episodes of scientific discovery begin with an individual with the “skill, wit, or genius to recognize that something has gone wrong in ways that may prove consequential.”(p. 763) But, he hastens to add, “anomalies do not emerge from the normal course of scientific research

¹ The author wishes to thank Branislav Slanchev and Laurie Clark for thoughtful comments and useful suggestions on an earlier draft of this paper.

² The fad around “clever identification strategies” is but the most recent instantiation of this phenomenon.

until both instruments and concepts have developed sufficiently to make their emergence likely and to make the anomaly which results recognizable as a violation of expectation.” (p. 763)

In the parlance of social media, scientific discovery begins with a “WTF” moment. Scientific discovery begins when a scholar observes something contrary to expectations and recognizes that this anomalous observation “may prove consequential.” Note that the motivating fact may be an observation about the world, but it may also be about what others have said about the world.³

But not just any surprise will do. Anyone who has ever parented a young child is familiar with the questions, born out of wonder, such as those that our children asked my partner and me: “why is the sky blue?,” “where does the sun go (at the end of the day),”? or, “if my brain controls my body, why do I have to go to the doctor to find out what’s wrong with me when I am sick”? Answers to all of these questions (assuming they are consistent with what scientists currently believe) are discoveries for the inquirer because they change what *they* know, but they do not lead to *scientific* discoveries unless they change what *we* know. The fields of optics, astronomy, and neuroscience have their respective answers to the questions above (although the last question is probably less settled than the other two).

So, questions often begin with surprise, but good research questions begin with well-informed surprise. If you alone are surprised by an observation, the answer to your “WTF

³ In the words of Branislav Slantchev, (personal communication) “theoretic innovation does not have to begin with an empirical observation but with a potential flaw in the logic, inconsistency of the assumptions, or an insight about a general claim (e.g., the impossibility results)”.

moment” is likely to be personally rewarding. If most well-informed observers are surprised by an observation, then an answer is likely to be socially, and, therefore, *scientifically* valuable.⁴

But sometimes, science proceeds when an individual recognizes that the answers embodied in what “we know” about a subject are not very good. For example, for millennia “we” knew that the answer to the question “where does the sun go” to be something like “the sun circles a stationary earth, so at a certain point each day it leaves our sight while shining on the other half of the planet only to return the next morning.” Eventually, however, scientists with “the skill, wit, or genius” to recognize the mounting anomalies created by models based on a geocentric view of the universe came to the conclusion that a better answer was needed. At first these better answers came in attempts to modify the heliocentric view with elaborate patches meant to explain away anomalous observations. In addition, to “skill, wit, and genius” it required a great deal of courage to challenge the existing view in a more fundamental fashion.

So, good questions come from knowing what “we” know. But they also come from thinking deeply about what we know and being sufficiently unsatisfied with bad answers to take the risk of thinking differently about a problem. As with all the arts, good science seems to come from individuals and groups that engage in a certain kind of practice. I would like to begin this essay by commenting on what I see as a common structure of many great contributions to political science and international relations. Specifically, I will put forward a list

⁴ It is fashionable in many top graduate political science programs for faculty to say that “substantive courses” are a waste of time and enterprising students should have an almost single-minded focus on methods training. It is also commonplace for professors to complain that their students are not adept at identifying interesting questions. I suspect that these phenomena are not unrelated.

of five questions that, when answered well, are likely to produce work that asks and answers interesting and important questions and gives us a reason to be confident in those answers. In the second half of the essay I will ruminate on the kind of practice that I expect to lead to good question asking and good answer giving.

Five Questions

When I was in graduate school, one of my professors, D. Michael Shafer, taught me how to read. He did so by encouraging me to employ a template he created so students could record the key parts of what they read: “What is the dependent variable?” “What are the independent variables?” “What is the logic that ties them together? Etc.” I found this enormously helpful in getting through the ridiculous amount of reading required in my graduate classes. When I began teaching I shared this list with my students and over the years I have refined it for various reasons. I have come to believe that this list of questions is useful not just in focusing our reading efforts, but also in our research efforts. If you ask what the author’s answer to each of the following questions is, you will have a good summary of most articles or books in our discipline.⁵ If you ask whether the author has a good answer to each of the questions, you will have a good critique of the paper in question. And if you are impatient with any bad answers provided by the author, and develop better ones, you will be on your way to making your own contribution to the literature. Consequently, I have come to believe that these questions can also serve as an excellent guide when designing a research project. If you have good answers to these five questions (and at least one of these answers is an

⁵ The questions would have to be adapted to serve this purpose for literature reviews, methods papers, and purely theoretical papers.

improvement over existing work), you will have a good paper, dissertation, or book. These questions also correspond to the organization of the modal paper in our discipline: “Introduction”, “Literature Review”, “Theory”, “Research Design”, and “Findings”.

It is important to add that research questions need not be generated by reading. They can just as easily, and perhaps more profoundly, be provoked by our interaction and observation of the social world. We might observe behavior and say “why does that happen”? It is good practice to offer one’s tentative answer to such a question unencumbered by “the literature.” But it is imprudent to spend very much time on such activity before evaluating existing answers to your question.

Question 1: What do I wish to explain? (The Introduction)

Following Kuhn’s description of scientific revolutions, most good work begins with a puzzling observation. Beginning with observation is important because good readers would like to be convinced that the phenomenon you are explaining actually occurs (though it is frequently fruitful to engage in thought experiments about things that have not occurred). This step is by no means trivial and considerable methodological sophistication may be necessary to accurately describe the real world events or, better still -patterns of events - which you wish to explain.

Samuel Huntington’s classic *Political Order in Changing Societies* (1968) seeks to explain the rising political instability he observed around the world. As evidence of this rising instability, on page four of this 462 page book, the author presents U.S. Department of Defense data showing that the number of nations around the world experiencing military conflicts of

various types rose almost monotonically from 34 in 1958 to 57 in 1965 (Table 1.1). This is a dramatic increase: in less than a decade the number of conflicts nearly doubled! The problem, however, is that, as a result of decolonization, the number of independent countries in the world also grew rapidly during this period. If one takes Huntington's numbers and divides them by the number of independent countries in each year (as a measure of the opportunity for military conflict), the *relative frequency* of military conflict actually declined over this period. Since military conflict was just one proxy for political instability, it is entirely possible that political instability actually increased during the observed period. But if you believe that

TABLE 1.1. Military Conflicts, 1958-1965

	1958	1959	1960	1961	1962	1963	1964	1965
Prolonged, irregular or guerrilla insurgency	28	31	30	31	34	41	43	42
Brief revolts, coups, uprisings	4	4	11	6	9	15	9	10
Overt, militarily conventional wars	2	1	1	6	4	3	4	5
Total	34	36	42	43	47	59	56	57

Source: U.S. Department of Defense.

the relative frequency of conflict is a better indicator of political instability than the raw frequency, you would be justified in wondering if the phenomenon explained in the subsequent four hundred or so pages actually occurred.

The first order of business, therefore, in demonstrating that something that may prove consequential has happened; is to demonstrate that *that* thing has happened. This crucial task

is often best accomplished with the presentation of clearly presented, well thought out, descriptive evidence. While this often requires a fair amount of methodological skill, sometimes it simply requires numeracy – which, unfortunately, is often in short supply. Effectively presenting evidence for one’s explanandum is, perhaps, best described in the breach. For example, you can read newspaper headlines on almost a daily basis that purport to capture some important change in the world that is, in fact, not supported by the text of the accompanying article. Would that it were the case that these mistakes were rare in academic work.

One common mistake is to make a claim about inter-temporal change in a variable by citing only current values of that variable. “Tenure-track jobs are disappearing” reads the title of an article, but the article makes no reference to the number of such jobs that were available in the past. How do we know that change has occurred? A related issue that requires a bit more methodological skill to avoid is to point out a difference between the values of a few recent values of a variable from preceding values and claim that they are evidence of a new trend without comparing the new observations with a long enough trend of data to determine whether they represent a meaningful deviation from the trend or, as is often the case, just typical variation within the trend.

Another common error is what might be called “the denominator problem” – the failure to choose a denominator that would transform the data into a variable appropriate for the conceptual comparison relevant to the discussion at hand. We already saw an example of this when Huntington confused a trend in the raw frequency of a variable for a trend in the relative frequency of the data, which I argued would have been more appropriate. But it is also possible

that the raw number is what most interests us – in which case we should not be distracted by an apparently related ratio. To return to the “disappearing tenure track jobs” problem we often hear about in the popular press, in the rare instances where inter-temporal data is presented in an attempt to establish this trend, the quantity presented is typically the ratio of tenure track jobs to the total number college teaching jobs. This is problematic because it is entirely possible for the *share* of tenure track jobs to be declining when the *number* of tenure track jobs is increasing (as has been the case in the United States for decades). And it is probably the latter number that is of interest to most readers (for example, current doctoral students hoping to forecast future demand for people with the credentials they are working hard to obtain).

Question 2: Why does it need to be explained? (The Literature Review)

Having explained that *this* thing has occurred, it is important for authors to demonstrate that a) this thing violates expectations in some way (i.e. “something has gone wrong”) and b) that this violation may “prove consequential.” In other words, in the words of Miles Davis, “so what?”

Once again, it might be easier to say what one should not do. I once attended a practice job talk where a smart, hard-working and, subsequently, very successful scholar, when pressed to say what he was trying to explain, said that he was trying to explain why a particular variable varies. Being less supportive than I should have been, I asked, “do you have a theory that leads us to expect this variable to be a constant?” Variables vary. It is even in their name.

Observing that variation, therefore, hardly constitutes a surprise. So if variation in a variable does not constitute a violation of expectations, what does?

As a comparative politics scholar, it pains me to say that I have attended many seminar talks over the last few decades, most given by successful and influential senior scholars, where the work in progress is motivated by an assertion that is some variant of the following “puzzle”:

theory Q claims that high levels of variable X should cause Y to happen, but in country i at time t , X was very high, and Y did not occur.

The problem with this “puzzle” is that once the misunderstanding it is based on is cleared up, it is no longer a puzzle. The misunderstanding is this: with very few exceptions (I cannot think of one) the empirical implications of social scientific theories are best treated as probabilistic (Lieberson 1991). Whether one traces the reasons to the intrinsically probabilistic nature of all human behavior deriving from human agency, the limitations of our understanding, the fact that most (all?) social phenomena have multiple, context-dependent causes, or the possibility of classification error (did Y occur or did it not? was X really high or low? and compared to what?) it is best to think of our hypotheses as probabilistic. Which means the most theory Q can claim is that “high levels of variable X should make Y *more likely* to happen.”

Consequently, the fact that Y did not occur in country i at time t , despite the fact that X was very high is not, at least to my ear, particularly puzzling. Unlikely events are expected to happen occasionally. Consequently, one cannot reasonably call a probabilistic conjecture into question with a single null case. Doing so is like being puzzled about one’s uncle who lived to a ripe old age despite being a heavy smoker. This is not puzzling because the best scientific evidence is that smoking increases the *likelihood* of cancer, not that it *always* leads to cancer. In contrast, it

would be surprising to find an entire sub-sample of the population that appears to be immune to the deleterious effects of smoking, or, that after controlling for income or education (or any other potential confound), smokers are not more prone to cancer than non-smokers. In sum, since our theories typically justify expectations about patterns of data, it takes observations about patterns of data, not discrete data points, to violate those expectations.

While recognizing a pattern in the data is often necessary for generating surprise, it is by no means sufficient. Going back to the many comparative politics seminars I have attended, be wary of the scholar who selects a small sample of observations and demonstrates that a widely corroborated empirical regularity, such as the incumbency advantage, the democratic peace, Gamson's Law, Duverger's Law, or the resource curse, "doesn't hold" in that sub-sample. Why? Because social behavior is probabilistic, so even highly predictive empirical models yield predictions with non-zero errors. As a result, one can always find a sub sample of data where the broader pattern does not hold. Take any "football shaped" scatter plot, such as the famous scatter plot shown in Figure 1.1.⁶ One can select out a sub-sample of cases, such as those in the ellipse, to suggest that the regression line is flat or even negative even though there is clearly a positive relationship in the sample on the whole.

⁶ The scatter plot is based on data from an example from Pearson and Lee's (1903) examination of the correlation between the adult heights of fathers and sons.

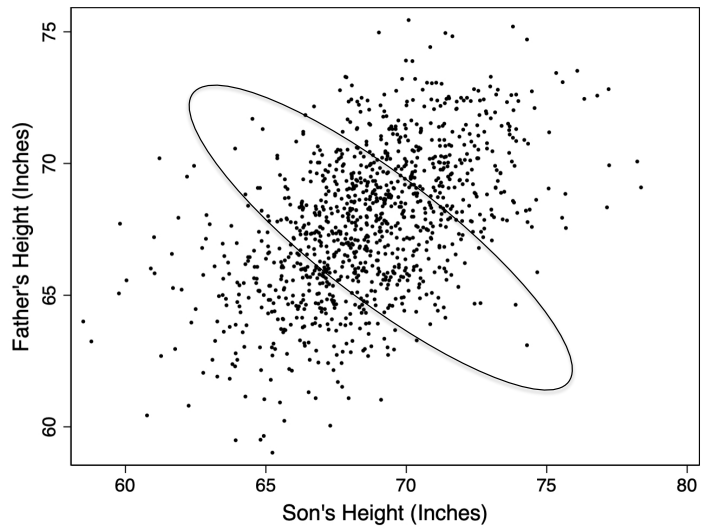


Figure 1.1 Relationship Between the Height of Fathers and Sons (Data Source: Freedman, Pisani, and Purves (2007) added random noise to data from Pearson and Lee (1903) who only had data to nearest inch,) <http://myweb.uiowa.edu/pbreheny/data/pearson.html>

Recall that I said, “be wary” of a scholar who motivates their study with a sub-sample of cases that appear to run contrary to a well-corroborated set of expectations. But I would not encourage you to dismiss such a scholar. It is, for example, entirely appropriate to show that there are boundary conditions on even the most well-corroborated empirical regularities. But the mere existence of such a sub-sample does not constitute a puzzle until one can convince the reader that the sub-sample constitutes a comprehensible category and is not just the result of felicitous (from the standpoint of the author seeking something to write about) case selection. Further, if one does take as their project the task of explaining why a well-corroborated regularity does not apply to a particular sub-sample, it is incumbent upon them to develop an explanation for why the sub-sample is different that yields predictions other than the fact that the sub-sample is different. Otherwise, they are engaged in both post-hoc and ad-hoc reasoning.

Yet another problem can arise when one generates their research project by gazing at a scatter plot. Many will look at a like Figure 1.1 after estimating a regression line and be disturbed that so many observations fall far from the regression line. It is okay to want the model to fit the data well, but given the probabilistic, multi-causal nature of our hypotheses, it is not puzzling that some observations fall far from the regression line. My father was six feet tall, while I, ahem, am not. That is not surprising because other factors enter into height at adulthood other than my genetic inheritance from my father – diet and contributions from my mother’s genetic make-up come to mind. Being puzzled in this way is a slightly more sophisticated version of the “if X is high in country i at time t , why do we not observe Y ” problem. Both methods are frequently used to justify the claim that “existing explanations are

incomplete.” The problem is that any explanation the author comes up with is likely to be susceptible to the same criticism.

I want to be clear, there is nothing wrong with being unsatisfied with explanations that do not fit the data well. However, if the only result of pointing out observations that fall off the regression line is a new model that marginally increases measures of goodness of fit, do not be surprised if readers fail to see this as “consequential.” *Ceteris paribus*, papers that are motivated by the identification of unclear, misleading, or incorrect understandings in the existing literature are more consequential than those that point to merely “incomplete” understandings because the former causes us to revise (that is to “look at again”) rather than merely supplement our current understanding.

So far, we have been seeking to identify violations of expectations that are consequential for our understanding of the world, but one might also place a priority on consequences that are more practical. One way of asking the “so what” question is to ask, “if you were successful in explaining your anomalous observation, how would the world be different?” Unless one is entirely naïve, this is a very tough question to answer. But since most of us became political scientists and international relations scholars because we wanted to make the world a better place, it is still worthwhile. One reason to think about the “normative” implications of the questions we ask is that an even passing familiarity with the literature in political science and international relations is enough to unearth a seemingly endless supply of unclear, misleading, or incorrect understandings. In light of this, it is not unreasonable to try to tackle first those that are tied to issues we care deeply about.

Nobel laureate Robert Lucas said “once you start thinking about economic growth, it is hard to think about anything else.” I suspect that is because it is not hard to see the real world, stick to your ribs, consequences of economic growth. Likewise, immigration, political violence, economic inequality, government corruption, racial and ethnic discrimination, financial instability, authoritarianism, gender bias, illiteracy, failing schools, or a host of other policy issues are of interest because of their impact on matters of justice and human well-being. Explaining observations that violate our expectations can be quite consequential when doing so sheds light on these and other social problems.

Marx’s last and most famous theses on Feuerbach is that “the philosophers have only *interpreted* the world, in various ways. The point, however is to *change* it” and it is interesting that it is etched on his tomb despite having never been published while he was alive.⁸ It captures the frustration of many scholars who would like to “make a difference.” It certainly captured my romantic heart when I first read it as a young man (not much younger than Marx was when he wrote it) at the start of graduate school. But I was not in graduate school long before I realized the complexity of “interpreting” the world and the dangers that could result if one sought to change the world without having interpreted it correctly. Understanding the world is a prerequisite for changing it in a responsible manner.

While it is desirable, perhaps even noble, bridging the gap between studying the world the way it is and using this information to improve social conditions is difficult - particularly when people, and, therefore, politics are involved. One problem is that if social ills have political roots even accurate explanations of their causes are likely to be insufficient for

⁸ Marx (1888).

mitigating them. One reason for this is the fact that the hallmark of politics is conflicting values. Explaining to prisoners confronted with plea deals that reward them for incriminating each other that they collectively benefit by keeping mum will not solve the prisoner's dilemma because they will still have individual incentives to rat on their co-conspirators.⁹

So, while understanding the world may be a necessary condition for (responsibly) changing it, it is not likely to be sufficient. And, conversely, changing the world can make it a lot harder to understand. One of the things that makes social science difficult is that the entities we study can read what we write and change their behavior in ways that make our models less predictively accurate.¹⁰

Something like this may have been at work in the writings of Marx. The phrase "workers of all lands, unite!" also appears on Marx's tomb. In contrast to his theses on Feuerbach, this phrase was published during his life-time. Three years after bemoaning the irrelevance of prior philosophers Marx and Engels closed one of the most influential political pamphlets ever written with it.¹¹ In an 1890 appendix to *The Communist Manifesto* Engels admits that few heeded the call in 1848 but suggests many eventually did so over time, including those who were organizing in support of the eight-hour workday 1890. It is not unreasonable to suggest that Marx's analysis of an internal logic to capitalism (that the

⁹ In contrast, if the only problem is a co-ordination problem then the mere dissemination of information is likely to be sufficient. But such problems are about as political as getting drivers to stay on their side of the road.

¹⁰ Though sometimes this works in the opposite direction. For example, experiments have shown that students who take economics classes behave much less cooperatively, and, therefore, more in line with the models learned in these courses.

¹¹ *The Communist Manifesto* (Marx and Engels, 1996) had little immediate impact on embryonic socialist movements, but its long run influence is undeniable.

inexorable immiseration of the proletariat would lead to revolution) helped fuel the formation of labor unions and the creation of social programs that improved the material conditions of workers. But in doing so, this made them less revolutionary – thereby reducing the probability of the revolution he predicted.

Another example of how it is hard to have both influence in the real world and predictive accuracy comes from the recent literature on “the happiness curve” – the robust empirical regularity that reported life satisfaction tends to decline when people are in their forties and rise consistently starting in their early fifties (Rauch 2018). One explanation for this empirical regularity is that because human psychology is biased towards overly optimistic forecasts, young people over-estimate how much their lives will improve in their thirties and forties. This results in disappointment during their middle years even if individuals’ lives have improved considerably, but not as much as they expected. This disappointment also leads people to update their expectations and make grim forecasts for the future. Consequently, when life in their fifties, sixties, and beyond turns out to be not as bad as expected, they report high levels of life satisfaction. If this process is truly at work, people who read this literature might be inclined to make more realistic predictions about future life satisfaction. If they did so in large numbers, the “happiness curve” could disappear.

Notice that to the extent that Marx changed history it may have been in ways that frustrated both his predictive accuracy and his social desires (for revolution) but if happiness researchers turn out to have the same degree of impact on society they might be perfectly willing to trade predictive accuracy for tangible improvements in people’s life satisfaction.

In sum, we would like to answer questions that, when answered, would prove consequential. These consequences can be either for the way we think about the world, or the way people behave. While, all else equal, we would like our research to lead to improvements in human well-being, the strategic nature of politics means that even when we provide good answers to questions that are important to us it may not lead directly to improvements in social outcomes. That is not to suggest we should stop trying.

Question 3: What is the explanation? (Theory)

A good explanation will take an observation that is sufficiently surprising that it justifies your study, and turn it into something that, in retrospect, should have been expected all along. In what remains one of the few books I know of that attempts to teach people how to explain things, the authors of *An Introduction to Models in the Social Sciences* (Lave and March, 1975) describe explanation as a process in which one imagines a prior world such that, if it existed, the surprising fact(s) would have been expected. Technically, any set of statements that logically imply the occurrence of the anomalous observation constitute *an* explanation. But good explanations have additional attributes, and we would like to produce the *best* explanation. A satisfying explanation will give the reader an understanding of the process or mechanism that is likely to produce the previously anomalous observation. Readers want to know how surprising events came about, and explanations should tell them. Good explanations are efficient – the ratio of things they explain (implications) to things they require you to believe (assumptions) is high.

There is an optimal degree of novelty to an explanation. An explanation should be interesting, yet sound. By “interesting” I mean that an explanation should cause us to see the world in a new way. By “sound” I mean an explanation should fit in with other things we know about the world. An explanation that causes us to see everything in a new way is likely to be wrong. An explanation that does not require us to change our mind at all is probably just a corollary of things we already knew (and, by extension, our motivating puzzle must not have been much of a puzzle).

Finally, explanations must be logically consistent. I have had empirically minded political scientists and international relations scholars tell me that formal theory is not important because they are sophisticated enough to live with theories that contain contradictions. This is nonsense. It can be shown with elementary logic that anything follows a contradiction. Consequently, if your theory contains a contradiction, anything can be said to follow from it. As a result, a contradictory theory rules nothing out and, therefore, no amount of empirical information will be sufficient to falsify it. Since potential falsification is the hallmark of science, a theory that contains a contradiction, therefore, is not a scientific theory.¹²

One way to increase the likelihood that your explanation is logically consistent is to try to capture it with a formal model. Formal models allow us to demonstrate that our explanation’s conclusions follow from its assumptions - most importantly, that our previously puzzling observation is not surprising in light of the world that our explanation posits. Also, by

¹² In the possibly apocryphal words of theoretical physicist Wolfgang Pauli, it is “not even wrong.”

making the assumptions of our explanation explicit we are more likely to notice if they contradict each other.

While these benefits of formalization are undeniable, it does not follow that every explanation should be formalized. I typically encourage my students to first articulate their explanations as a story that reveals a process that produces the previously unexpected observation. Formalization is only necessary when one hears such a story and asks, “why would people do that?” or, equivalently, “That doesn’t sound like an equilibrium,” or, “isn’t there a tension between this part of the story and that part of the story?” When one is confronted with such questions, a good formal model can often provide answers. Thus, I tell my students to learn how to write down formal models not because they will always need one, but like fire insurance, they are always at risk of needing one.

Another reason to begin with an informal statement of one’s theory is to avoid the trap of thinking that a game theoretic model will generate a theory for us. Formal models help us interrogate certain aspects of our theory, they do not produce the theory for us. We must begin with some theoretical intuition about what explains the phenomenon in question before we can begin to model the process.

Question 4: If the explanation is true, what else should we observe? (Research Design)

If you offer a view of a theoretical world that has the previously puzzling observation as one of its implications, you have offered an explanation. And while there are various ways to evaluate that explanation, to be scientific, your answer to your original question must provide an answer to the following question: “if your explanation is correct, what else ought to be

true?” Good scientific explanations provide lots of answers to this question. If your explanation only implies the facts that you set out to explain, then there is no way to empirically evaluate your answer. You cannot use the fact that democracies seldom fight each other, or the fact that there is a lot of corruption in presidential democracies to evaluate your explanation of these things because it was those facts that led you to develop your explanation in the first place.

This part of the research process is a stumbling block for many researchers when they are attracted to a subject, rather than a question. I once had a student who visited Brazil and was shocked by the level of corruption in the government there and developed an explanation that pointed to aspects of the large district magnitude proportional representation electoral system as a cause. The student was surprised when I said I thought the argument had merits, but that returning to Brazil to collect data was not a promising avenue for evaluating the argument: we already knew that Brazil fit the argument! Perhaps data on corruption levels in countries with different electoral laws (such as the United States) would be more useful. The student, however, responded that he did not want to study corruption in other countries, after all he was interested in Brazil!

A similar problem is found in a very famous book by Theda Skocpol, *States and Social Revolutions* (1978). In it, the author wishes to explain the occurrence of social revolutions and she argued that her subject dictated her empirical strategy. Given her definition, there are only five historical cases of social revolution. She argued that as a consequence of this fact, structured focused comparison (specifically, Mill’s Method of Agreement) was the only possible method for evaluating her explanation. That is not true.

The chief problem here is that if an explanation for a set of rare events only has implications about those rare events, the author does not have a data problem, they have a theory problem. If an explanation for global warming only predicts the general rise in the temperature that motivated the explanation, then it is not a very useful explanation. Cosmologists have offered explanations for the creation of the universe, but they do not choose their methodology for evaluating their explanations based on the fact that the object of their study only happened once. Instead, they ask, “if my explanation for this unique event is correct, what else ought to be true?” They then think about how best to carefully observe the implications of their argument.

The goal of empirical research, therefore, should be to examine as many implications of one’s explanation as possible. Because many, many scholars restrict their attention to the empirical puzzle that motivated their study to begin with, many important papers can be written by simply asking of existing explanations, “if this argument is true, what else ought we observe?”

One reason why scholars often restrict their attention to the data that generated the question is that it can often take considerable creativity to think about the implications of an explanation. There is no cookbook-like approach that can be applied that will automatically reveal to the scholar that seemingly unrelated events might be instantiations of a single social process. But one practice Lave and March recommend is to try to see your answer to a particular question as related to a more general process.

For example, in her critical review of Skocpol’s book, Barbara Geddes (2003) suggests that one element of Skocpol’s explanation of rare social revolutions had implications for the

occurrence of peasant revolts. Geddes suggest that a statistical model examining the conditions under which peasant revolts do and do not occur would, therefore, be useful in evaluating the empirical relevance of Skocpol's explanation of social revolutions.

Notice, that when we ask "what else ought to be true" we separate the question of "what is the authors explanandum?" from "what is the author's "dependent variable?" The explanandum is a statement of what the author develops a theory to explain. The "dependent variable" is the endogenous variable in a model testing one or more of the implications of the author's theory. There are times when these might be the same, but there is no reason to assume they will be. In fact, when they are, we should wonder if the author is engaged in post-hoc reasoning – "have they observed the dependent variable and its covariates and constructed a causal story after the fact?" Doing so would constitute a "test" of the theory only to the extent that the lion's share of the observations could be thought to have been appreciably different from those that were observed before the theory's formulation. Conversely, a theory that produces a lot of novel implications helps assuage the reader's suspicion that the author is merely engaged in a curve-fitting exercise.

In sum, it is typically more helpful to think of empirical work as testing the implications of a theory, rather than testing the theory directly. One reason this is true is that testing the theory directly can easily descend into more or less complicated version of curve-fitting and post-hoc reasoning. Instead, spend time thinking about the implications of your explanation for observations other than those that motivated your question in the first place. The more varied those implications the better, because it is only those observations that are made after the construction of your theory that run the risk of being false and therefore actually constitute an

empirical check on your explanation. And remember: if your theory only has implications for a set of events too small to use standard inferential tools to evaluate, you do not have a data problem - you have a theory problem.

Question 5: Do we observe the implications of our explanation? (Findings)

Determining if evidence is consistent with one's theoretical expectations is the primary focus of research methodology and, so, is the central focus of the remainder of this volume. Here I will merely stress the following: many, many studies present, often in dizzying detail, reams of information that is either irrelevant to or inconsistent with theoretical expectations. Typically, however, it is presented in a manner that suggests that this information confirms the author's expectations. Distinguishing when this is the case is a large part of what is meant by learning to read critically.

As I said, all of the collective wisdom of research methodologists is relevant for becoming a critical reader and producer of knowledge but I will focus on one admonition: present clear estimates of the quantities of interest as well as a statement about the degree of confidence one has in those estimates.¹³ There are a few ways in which this admonition is frequently violated, and I would like to briefly draw your attention to them.

At least in the social scientific papers I read, explanations typically produce claims about the association between variables. Even when one is engaged in what looks like a descriptive exercise, like Huntington's attempt to demonstrate rising political instability, one is engaged in demonstrating that variables are related to each other in a particular way. If one wants to

¹³ King, Keohane, and Verba (1994).

demonstrate that a phenomenon is changing over time, one must look at the relationship between that variable and time. If one wants to demonstrate that a particular behavior or attitude is more prevalent in some places or among some groups, one must look at the relationship between that variable and group membership or spatial location. Consequently, most of our empirical claims are about the relationship between variables. In a linear model we think of this quantity of interest as a slope coefficient, so I will use that terminology here, though the term “derivative” might be even more appropriate.

A common way in which scholars become distracted from presenting the quantity of interest is by presenting something other than an estimate of a slope, when that is the quantity they are concerned with. For example, it has become common for scholars to plot the predicted probabilities from a logit model on the y-axis with some variable of interest on the x-axis when the quantity of interest is the association between a change in that predicted probability and a meaningful change in some variable of interest. The problem with doing so is that it requires the reader to infer the slope of that relationship from the picture. While it is true that slopes are not constant in non-linear models such as logit, and, therefore the quantity of interest does not reduce to a single number, it would be better to plot the marginal effect of the variable of interest across a meaningful set of values of that variable of interest.¹⁴ Adding confidence intervals around the predicted probability does not help because that tells the reader if the predicted probability is significantly different from zero, which is typically not the hypothesis being tested.

¹⁴ In the language of calculus: if the quantity of interest is dy/dx , then plot dy/dx against x , not y against x . The former tells the reader what they need to know. The latter makes the reader try to infer what they need to know from the picture.

For example, Hellwig, Ringsmuth, and Freeman (2008) present the graphs in Figure 1.2 as evidence in that the propensity for citizens to believe governments have little room to maneuver policy in a globalized economy. Each panel plots the predicted probability (and 90 % confidence intervals) that a survey respondent said they did not believe the U.S. government retains the “room to maneuver” policy against the respondent’s partisanship. The authors interpret the apparent difference between the slope of the plots in the left hand panel from the right hand panel as evidence that partisanship has an effect on respondent beliefs among respondents with College Degrees (panel a) but not with High School Degrees or Less (panel b) and among respondents above the age of fifty-nine (panel c) but not below the age of forty (panel d). But what is the basis of this conclusion? The slopes on the right clearly look to be close to zero and, in comparison, the slopes on the left appear to be positive. But we are offered neither an estimate of the slopes for any degree of partisanship, nor an estimate of our uncertainty about that estimate. We can try to calculate the slope at different points on the line by estimating the “rise over run” and we can kind of compare that estimate with the uncertainty implied by the error bars, but why make the reader construct a t-test from the picture rather than present that information for the reader by plotting marginal effects with their associated confidence intervals? Neither do the authors provide any evidence whether the slopes in the left-hand panels are different from the slopes in the right hand panels. As a consequence, these pictures, and ones like them that appear frequently in the literature, provide almost no *quantitative* evidence about the quantity of interest (under what conditions, if any, a change in partisanship associated with a change in citizen beliefs about the government’s “room to maneuver”).

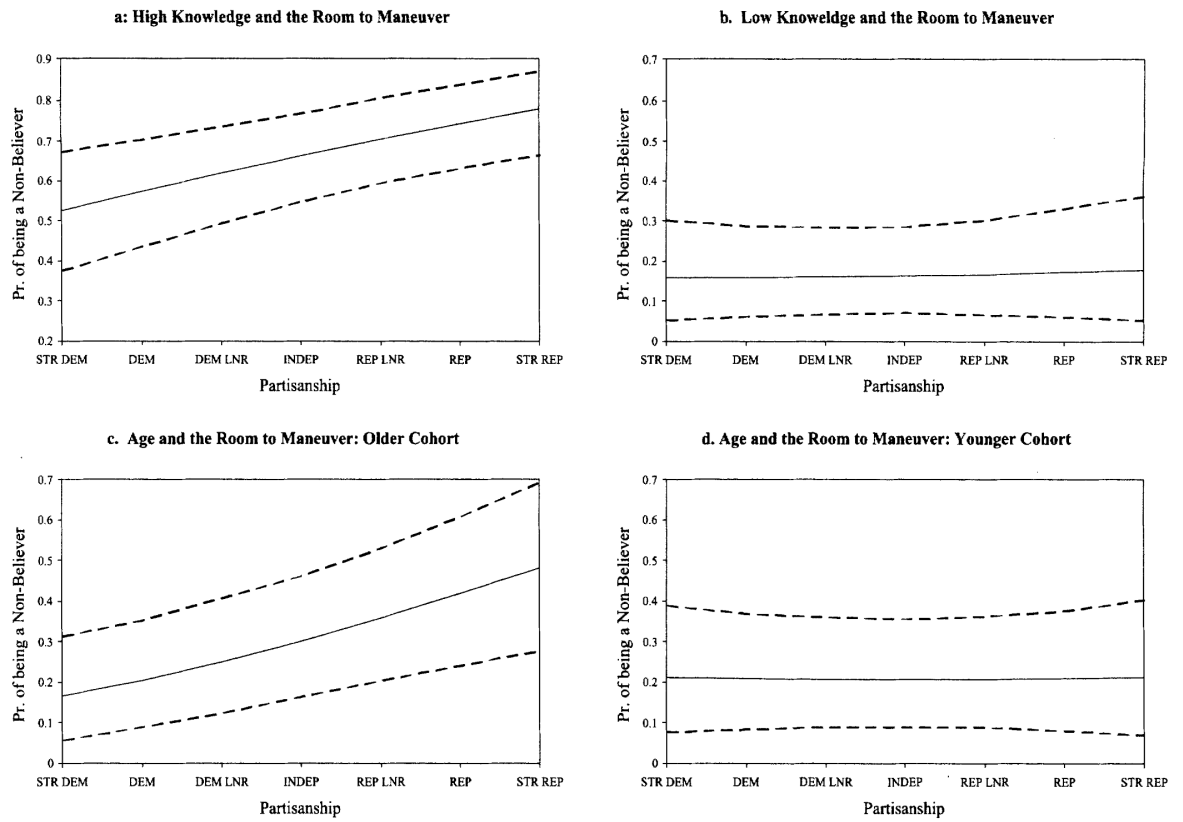


Figure 1.2 Partisanship and Beliefs about 'Room to Maneuver: The Conditional effects of Knowledge and Age. Source Hellwig, Ringsmuth, and Freeman (2008, Figure 2, p. 875.)

Another common way of obscuring the quantity of interest is in presenting “marginal effects” that are not marginal. It is common place for authors to say things like “to gain some substantive understanding of these results, I note that a one standard deviation change in X is associated with a 0.056 change in Y.” The problem with this is that there is nothing typical or representative about a standard deviation – in data approximating a normal distribution about two-thirds of all observations will be less than a standard deviation away from the mean. As a consequence, a change of a standard deviation in the variable of interest is not a particularly meaningful counterfactual to consider. This is particularly true where this practice is most frequently found – when interpreting the results of a non-linear model. Under this circumstance, the marginal effect of a variable is extremely sensitive to where it is being evaluated. The slope described by a “marginal effect” the size of a standard deviation is likely to be very far from the slope of any estimated marginal effect within this interval. Another reason why this is not a particularly useful counterfactual comparison is that marginal effects are interpreted under a *ceteris paribus* clause where other factors are held constant – something which is not likely to be approximated in the real world when the variable of interest experiences an unusually large change the size of a standard deviation.¹⁵

Another common way scholars present information that is not the quantity of interest is when they have a hypothesis that is conditional in nature and either present results from an unconditional model, or, equally common, estimate a conditional model but go on to interpret some of its results as if they were unconditional.¹⁶

¹⁵ See King and Zeng (2006) on “The Dangers of Extreme Counterfactuals.”

¹⁶ See Brambor, Clark and Golder (2006) or Kam and Franzese (2007) for a fuller discussion.

Summary

My claim, up to this point, is that a paper, book, or dissertation that has good answers to the five questions above will be a useful paper, book, or dissertation. It does not follow that a paper, book, or dissertation must have an *innovative* answer to all five of those questions. Progress can be made as long as one of the answers is better than existing answers and none are worse.

Which questions are “most important” and, therefore, which ones should be the focus of your efforts to innovate? It is hard to say. Though I believe that it is probably not best to try to explain something that no one has explained before. This is an important point. I have had many graduate students inform me gloomily that someone has beaten them to their “question.” My standard reaction is to say, “well, I doubt they have come up with the definitive answer, so what are you worried about?” Since any question worth asking is likely to be difficult to answer, it is highly unlikely that another scholar is likely to “beat you to the punch” and state the “last word” on a subject. Indeed, if you are asking a question that no one else has asked, it should give you pause. Maybe it is not a very interesting question: or maybe there is something about asking the question in that way that led other scholars to believe productive answers were not forthcoming. That said, the mere fact that other smart people have asked the question does not mean it is a great idea for you to try to answer it.

Graduate students are told that they need to make an original contribution which leads them to believe that they must ask a question that has never been asked, or at least, never been answered before. That is not true. Rather, an “original contribution” requires only that the student provide a better answer to at least one of the questions mentioned above. So, if a

student at the prospectus stage is going to attempt to offer a novel explanation, then part of their answer to question 2 should contain a statement about what they bring to the table that might allow them to make progress where others have failed. What theoretical insight, methodological advantage, or historical knowledge puts the author in a position to simultaneously recognize that “things have gone wrong” with existing explanations and offer a solution that pushes the field in a promising direction?

Since “theoretical innovation” is often thought to be the most prized contribution a political scientist can make, scholars often believe that a good paper should offer a novel explanation. I believe this comes, in part, from physics envy combined with the notion that theoretical physicists have a higher status than experimentalists. I believe the idea that every important contribution must contain a theoretical innovation has greatly hampered the progress of our discipline. How is the accumulation of knowledge possible if every time a scholar puts pen to paper they have to offer a new explanation? Given frequently imperfect research designs and flawed empirical methods, I often think the opposite is true. We might be tempted to declare a moratorium on the development of new explanations until the discipline has reached consensus about empirical tests of the implications of existing explanations. As my critique of Huntington suggests, if we do not get at least some of the empirics right, how do we even know if our observations violate current theoretical expectations enough to warrant new explanations? One reason to resist such a temptation is that new theories do more than explain anomalies. For one example, they also address conceptual and logical problems with existing explanations.

Practices that encourage good question asking

Following Kuhn's line of reasoning above, it is worth asking what is likely to promote "the skill, wit, and genius" capable of recognizing when things have "gone wrong in ways that may prove consequential." Of Kuhn's three desiderata, "skill" seems the least constrained by natural ability and, therefore, most responsive to the environments we create. While artistic creation involves many aspects, a degree of craftsmanship is typically involved and craftsmanship is derived largely from practice. Extensive training in game theory and statistics is now commonplace in most graduate (and some undergraduate) programs in political science and international relations and this is what is typically thought of when scholars evaluate the "skills" of job applicants. These skills are important because without them, scholars might ask questions based on faulty reasoning based on formal or informal fallacies such as the ecological fallacy, ad hominem attacks, hasty generalization, confusing correlation with causation, ignoring strategy induced selection effects, and failing to recognize the presence of confounds.

But while methods training is extremely helpful, it is not sufficient to produce scholars who ask and answer interesting questions. The problem sets typically assigned in quantitative methods and formal theory classes do help build the skills necessary to execute sophisticated research. Just as playing scales and arpeggios build the techniques necessary to execute sophisticated music. But there is more to training a musician than playing scales and arpeggios because as important as scales and arpeggios are, they are not music. I have heard musicians criticized for having sufficient technique that they "know how to say things on their instruments, but they do not seem to have anything to say." The analogous criticism is frequently leveled at newly trained political scientists and international relations scholars.

So, what is to be done? To play good music, students have to listen to good music and they have to have a lot of experience making good music. Most graduate programs provide students with the equivalent of listening to music. When I was a newly minted Ph.D. I heard Bruce Bueno de Mesquita give a lecture at the Hoover Summer Program in Game Theory and International Politics at Stanford University. He built a game theoretic model based on the assumptions of hegemonic stability theory – seemingly on the fly based on comments shouted out by my classmates. I had an epiphany. Of course, if developing social scientific explanations is an art, then it must be taught as the arts are taught! I was watching the master at the easel – engaged in the very craft I was trying to learn. It suddenly occurred to me that much of my graduate training amounted to the equivalent of sitting in a room listening to recordings of music, then when it was time to write my dissertation it was as if a door had been flung open, I was handed an instrument I had never played (I imagined a cello) and pushed out onto a stage where I was expected to perform. Most graduate programs in political science teach people the equivalent of playing scales in methods classes and music history or appreciation in substantive classes and are left to figure out on their own how to put this together to make music.

The missing piece in most of our graduate education is what musicians call “etudes.” These are exercises designed to be music-like (so students can begin to think about interpretation and expression) but are artificially designed to allow for a degree of repetition of particular techniques (articulation, vibrato, dexterity) that allows those skills necessary for musical expression to seep into the student’s muscle memory. Many doctoral programs emphasize that students should write publishable papers, but I believe that success is unlikely if

this is attempted before students have engaged in many repeated attempts to explain things or think about what observations are implied by their explanations. Students need to practice asking and answering the five questions outlined above and writing a single paper in each seminar does not give them the “reps” to develop muscle memory. Virtually no skill worthy of the name can be developed after a dozen or so attempts.

Consequently, I have argued that problem sets in “substantive classes” can help students become proficient at asking and answering the questions that will make for innovative research. An analogy to the visual arts might be useful. When students are learning to draw, they are not handed a blank sheet of paper and told to “think of something interesting to draw, that no one else has drawn.” Rather, a bowl of fruit, or perhaps a wooden model of a human figure is placed on a table. Then, everyone in the class draws the same thing after receiving instruction from the instructor about how to do so. In contrast, many political science departments do the equivalent of handing their students a blank sheet of paper and are told to “draw something interesting.” Problem sets in substantive classes can be the equivalent of a bowl of fruit. The instructor can assign students to a question related to a particular research area. “Explain why X occurs under Z circumstances.” “If P explains Y, what else ought we observe?” “Why is Q an interesting question?” “Does Figure 2 count as confirming or disconfirming evidence for hypothesis 2, and why?”

Students need a lot of experience “making music” before they “have something to say.” If the analogy to the arts does not resonate with you, consider the following. Political science and international relations can take a lesson from the so-called “bench sciences” where students work on many projects as members of large teams before they are tasked with the

responsibility of deciding on the topic of the group's next project. Experience and repetition helps students learn what works and what does not.

While graduate pedagogy is important for stimulating creative question asking and answering, the broader climate and culture we create in our departments and research centers is equally important. In particular, it is extremely important to create an environment where it is safe to play with ideas and challenge orthodoxy. I once had a colleague who, while walking down the hall read a passage from a book that he thought was incorrect and loudly declared the author an "idiot." Creativity and risk taking is not encouraged by a culture that suggests that only stupid people say stupid things. Instead, it is important to create the idea that the smartest among us are capable of error and that there is a big difference between saying something that stupid and *being* stupid. To that end, I think it is extremely important for senior scholars to be transparent about the errors they have made. Young scholars need to learn that if they've made a mistake, they are in very good company and if the requirement for admission was never making a mistake the building would be empty.

While a culture of support for individual risk taking is vital to any scientific or artistic community, there is an optimal degree of individualism behind scientific discovery. If you don't read what everybody else reads and fail to train like everyone else trains, you will ask naïve questions that the rest of your community knows the answers to. But if you only read what every reads, and only train like everyone else trains, you are unlikely to experience that moment when you see something that has gone wrong that no one else sees.

Jazz bassist Scott LaFaro started playing the bass in 1954 when he was 19 years old and in the few short years before he was killed in a tragic car accident in 1961, he completely

changed the world's conception of what could be accomplished on a double bass and what role the instrument could play in a piano trio. Prior to playing the bass he had played the clarinet and saxophone for years and many have attributed his phenomenal technical prowess to the fact that he practiced the bass by playing *etudes* composed for the clarinet by Hyacinthe Klosé in the 19th century (LaFaro-Fernandez, 2009). The lesson LaFaro taught the world, in addition to the general benefits of inter-disciplinarily, was “if you want to sound like everyone else, practice like everyone else; but if you want to sound like no-one else, practice like no one else.”¹⁷

Just as there is an optimal degree of individuality that is likely to produce scholars with the skill, wit, and genius to determine when something has gone wrong in ways that may prove consequential, communities that strike the right balance between conformity and diversity are likely to encourage the habits that lead to scientific breakthrough.

On the one hand, it is important for a scientific community to share a commitment to the growth and dissemination of knowledge and a common understanding of the logic of inference and the standards of evidence. Without this shared understanding, criticism is likely to fall on deaf ears. But on the other hand it is important for a community to be as diverse and eclectic as possible. People from different cultural, class, linguistic, and religious backgrounds are likely to see the social world differently because they are likely to have had different experiences. These different experiences are likely to lead to diverse moral, political, and social intuitions that lead them to raise questions that a more homogeneous group might not (Page, 2007).

¹⁷ At the same time, nearly every innovative jazz musician learned their craft by memorizing performances of musicians that came before them.

In addition, diverse groups are less likely to fall prey to what I call “strategic confirmation bias.” Confirmation bias occurs when an individual embraces an idea uncritically because it conforms to their prior beliefs. When confirmation bias is at work, people are less likely to scrutinize the research practices that produced the claim in question. They are less likely to look for confounds, to ask about the details of data collection, or to think critically about either the micro-foundations or moral implications of a claim because the results confirm what they have long suspected about the world.

But strategic confirmation bias occurs when an individual is able to overcome first-order confirmation bias and think critically about the claim being made, but is deterred from voicing the criticism because they believe others are refraining from criticism as a result of confirmation bias. Under such circumstances, critically engaging the claim in public might signal to others that the critic does not share their beliefs on the matter.

Strategic confirmation bias is most likely to be a problem in communities where “everybody” shares particular beliefs. In such an environment, thinking critically about a result that confirms the community’s beliefs could result in ostracism, or at the very least, fewer dinner invitations. A community comprised of individuals from diverse educational, class, religious, and ideological backgrounds is less likely to produce the kind of monolithic views that encourage strategic confirmation bias. Individuals are more likely to say something when they see something wrong that may prove consequential because the set of taken for granted shared beliefs is likely to be smaller. Diversity is most likely to be helpful in this regard when the multiple dimensions of identity are relatively uncorrelated. If gender, race, or ideology are heavily correlated, then dissent on one dimension can be seen as defection on another. Thus,

in ideal circumstances communities would have as much within group diversity as between group diversity.¹⁸ Of course, diversity has to be sufficiently developed to give individuals confidence that speaking up under such circumstances will not simply confirm that one is an “outsider.” If a community promulgates the norm that in a multidimensional space we are all, on one dimension or another, outsiders, the cost of revealing that one “thinks differently” about something is likely to be less costly. The daunting thing about strategic confirmation bias is that it is mostly likely to occur around issues scholars feel passionate about. As a result, there is a danger that a research community will be least scientific about the matters that it cares most deeply about and most scientific about matters its participants view as largely inconsequential.

CONCLUSION

Good scientists ask interesting questions and are unsatisfied, even impatient, with bad answers. I have argued that most work in political science and international relations can be understood through the lens of five questions and that contributions can be made to the literature by improving on a research community’s answer to any of the five questions.

Since coming up with better answers to questions is as much art, as it is science, I have argued that the best way to train good social scientists is to learn from the way artists are trained. Musical and visual artists learn their crafts through structured repetitive practice. The implication of this insight for the social sciences is that scholars should be given materials to work with that allow them to engage in the daily practice of asking and answering the five

¹⁸ The connection between “intersectionality” and cross-cutting cleavages should be explored further.

questions outlined in the first section of the paper. I have suggested that the best way to encourage this is through the use of problem sets in our substantive courses. I have also hinted that there are great benefits to inter-disciplinarity. By bring habits, techniques, and insights that are normal in one discipline to a setting where they are rare, individuals are more likely to recognize when something has “gone wrong in ways that may prove consequential.” Finally, I have argued that diverse communities are more likely to produce good question askers, in part because they are less likely to fall prey to strategic confirmation bias.

References

Brambor, Thomas, William Roberts Clark and Matt Golder. 2006. "Understanding Interaction Models: Improving Empirical Analysis" *Political Analysis*. 14(1): 63-82.

Cindy D. Kam and Franzese, Robert J. Jr. 2007. *Modeling and Interpreting Interactive Hypotheses in Regression Analysis*. (Ann Arbor: University of Michigan Press).

Friedman, David, Robert Pisani and Roger Purves. 2007. (New York: W.W. Norton).

Geddes, Barbara. 2003. *Paradigms and Sand Castles: Theory Building and Research Design in Comparative Politics*.

Hellwig, Timothy T, Eve M. Ringsmuth, and John R. Freeman. 2008. "The America Public and the Room to Maneuver: Responsibility Attributions and Policy Efficacy in an Era of Globalization." *International Studies Quarterly* 52(4).

Huntington, Samuel. 1968. *Political Order in Changing Societies*. (New Haven: Yale University Press).

King, Gary, Robert O. Keohane, and Sidney Verba. 1994. (Princeton: Princeton University Press).

King, Gary and Langche Zeng. 2005. "The Dangers of Extreme Counterfactuals," *Political Analysis* 14:131-159.

Kuhn, Thomas S. 1962. "Historical Structure of Scientific Discovery," *Science* 136(3518):760-764.

Lafaro-Fernandez, Helene. 2009. *Jade Visions: The Life and Music of Scott LaFaro*. (Denton, Tx: University of North Texas Press)

Lave, Charles A. and James G. March. 1975. (New York: Harper and Row).

Lieberson. Stanley. 1991. "Small N's and Big Conclusions: An Examination of the Reasoning in Comparative Studies Based on a Small Number of Cases." *Social Forces* 70: 307-320.

Marx, Karl. 1888. "Theses on Feuerbach" in Friedrich Engels, editor, *Ludwig Feuerbach and the End of Classical German Philosophy*.

Marx, Karl and Friedrich Engels. 1996. *The Communist Manifesto*. (London: Pluto Press).

Page, Scott. E. 2007. *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies*. (Princeton: Princeton University Press).

Pearson, K. and Lee, A. (1903). On the laws of inheritance in man: I. Inheritance of physical characters. *Biometika*, 2(4), 357-462.

Popper, Sir Karl. 2003 [1959]. *The Logic of Scientific Discovery*. New York: Routledge.

Popper, Sir Karl. 1962. *Conjectures and Refutations: The Growth of Scientific Knowledge*. New York: Basic Books.

Rauch, Jonathan. 2018. *The Happiness Curve: Why Life Gets Better After 50*. (New York: St. Martin's Press).

Skocpol, Theda. 1979. *States and Social Revolutions: A Comparative Analysis of France, Russia, & China*. New York: Cambridge University Press.