# 10

# The Methodology of Presidential Research

GARY KING

THE ORIGINAL purpose of the paper this chapter was based on was to use the Presidency Research Conference's first-round papers—by John H. Aldrich, Erwin C. Hargrove, Karen M. Hult, Paul Light, and Richard Rose—as my "data." My given task was to analyze the literature ably reviewed by these authors and report what political methodology might have to say about presidency research. I focus in this chapter on the traditional presidency literature, emphasizing research on the president and the office. For the most part, I do not consider research on presidential selection, election, and voting behavior, which has been much more similar to other fields in American politics.

I am in an odd position in this task, since many of the various topics on which methodologists usually give advice do not apply here. In other fields, we are often asked how to estimate the parameters of a particular model with available data, how to develop a model from a more vaguely expressed theory, how to measure a concept more accurately, or how to avoid statistical biases when available data have a variety of specific limitations. Judging from the first-round papers, one might think that few presidency researchers pose questions like these. Although probably more has been written about the presidency than all other areas of American politics combined, most work in the field is not yet to the point where concepts are to be measured and theories tested systematically.

Yet, I do think that the presidency literature could benefit from some of the insights of political methodology. In this chapter, I address this state of affairs in presidency research in four increasingly specific ways. First, I directly discuss the *systematic* study of the American presidency, a subject about which presidency scholars often seem quite defensive, and I believe unnecessarily so. I argue here that the division between "rigor and relevance" made numerous times in these papers and in the literature is of limited value and that qualitative research can be as rigorous as quantitative research. Indeed, the rules of scientific inference apply to all areas of research equally, so we must hold qualitative presidency research to the same standards. Second, I discuss the explicit goal of the literature reviewed in every first-round presidency paper: increasing the richness of description and inclusiveness of theoretical perspectives. I take the position that this is not a productive direction for future research on the American presidency. Instead, we need much *less* inclusive and more *specific* theoretical concepts: a few very precise or even incorrect theories would serve the discipline much better.

Third, I argue that the famous $n = 1$ problem of presidency research is not at all specific to this literature and is indeed a perfectly general statement of the problem of causal inference. I also demonstrate in this section that the common practice of using the president as the unit of analysis is very unlikely to yield reliable inferences Although this problem is widely recognized in the presidency literature, its solution is not self-consciously understood even though much research does get around the problem. Finally, I make some very specific positive suggestions for the presidency literature by providing outlines of research projects that might help presidency scholars design research so they could learn what they desire to know and, at the same time, meet the standards of scientific inference espoused by political methodologists.

## *The Systematic Study of the American Presidency*

Presidency research is one of the last bastions of historical, nonquantitative research in American politics. In this section, I argue that the frequent questions about whether we should use systematic approaches are moot. An area of research is not systematic just because scholars use numbers or nonsystematic because they do not. Indeed, there is no inherent nonstylistic difference between quantitative and qualitative research—in the presidency or in any other area. To



the contrary, I have never heard anyone even argue that specific rules of inference are inapplicable in qualitative research; I see no reason, therefore, to treat qualitative research any differently than quantitative. Thus, we must apply these time-honored rules in order to evaluate and, where possible, improve presidency research.[1] It is from this perspective—using the rules of inference so clear in quantitative research to best understand the qualitative research in this field—that we can best evaluate inferences about the American presidency.

We should never permit a "balance" between rigor and relevance: we simply must demand both. Traditional presidency scholars are unquestionably correct in arguing that rigor without relevance is worthless: a good answer to a question no one cares about is of no value. Unfortunately, rigorous papers about irrelevant topics have appeared in the presidency literature, as elsewhere, so the criticism is to the point. By the same token, unconvincing analyses of important issues are equally suspect.

On the other hand, the presidency literature also contains countless books and articles with highly relevant arguments that are qualitative instead of quantitative. This is not necessarily troublesome at all, insofar as one should not automatically equate rigor and quantitative analysis. Although presidency researchers could bring methodologically sophisticated analyses to bear on relevant empirical questions in the presidency literature to a much greater extent then they currently do (see Edwards, 1983), much current work would not be improved by adding quantitative analysis. Rather, the standards for judging qualitative research must be clarified and scholars' qualitative work evaluated accordingly. The problem with this part of the presidency literature is *not* that scholars have failed to do something they did not wish to do, for we must judge these works on their own grounds. The problem is that the qualitative research that is done is not always as rigorous as it could be.

Precisely what about qualitative research in the traditional presidency literature could be improved? One could speak of many problems: incorrect inferences, misunderstandings of causality, measurement error, bias in case selection, spurious effects, and others. However, to some extent, all these problems afflict every kind of qualitative research—and much quantitative work as well. That these problems exist in the presidency literature does not make it unique or even unusual.

In my view, the signal problem with qualitative research on the presidency is its failure to appropriately judge the *uncertainty* of our inferences. One can make a valid inference in almost any situation, no matter how limited the evidence, but one should avoid forging policy recommendations out of thin data. The point is not that reliable inferences are impossible in qualitative research; rather, one should always report a reasonable estimate of the *degree of certainty* we have in each of our inferences.

For example, suppose you buy a house, move into a neighborhood, and shortly thereafter discover that ten of your neighbors have stomach cancer. A reasonable inference is that the water supply is badly contaminated. At this point, should you sell your house for a small fraction of its purchase price in order to move out of the neighborhood as quickly as possible?

You would probably want more evidence first, since the uncertainty of your original inference is too high for you to make such an important decision. (On the other hand, you might purchase bottled water for a while, a relatively low-cost decision.) The certainty of your inference would increase if you found that a river running through town looked polluted, and it would increase further if you found an industrial plant upstream dumping waste into the river. You might become very certain if you tested water in that river and found a known carcinogen present in high quantities. However, if you then found that drinking water for the town did not come from the river, the inference of a problem with the water supply would again become less certain. If a test of the town's water supply indicated no carcinogens, you might just be genuinely puzzled—no longer even willing to hazard a guess as to the cause of the high incidence of stomach cancer. Finally, if you discovered that your ten neighbors moved to town *with* the disease in order to be close to a major hospital that specializes in the treatment of these patients, the puzzle would be resolved, and you would dismiss your inference with a high level of certainty.

At each point in this ministudy, your estimate of the uncertainty of your inference about the water supply changed. If polled about your opinion at any point in this process, you would report both your inference—your best guess based on the available evidence—*and* an assessment of how good that best guess is. In no case would it make



much sense to report the inference without also reporting a judgment about the uncertainty of the inference. Indeed, reporting the inference with an incorrect assessment of its uncertainty can be as bad or worse than if your best guess were wrong.

In quantitative research, scholars routinely calculate point estimates and standard errors. We can think of these as best guesses and the uncertainty of these guesses, respectively. Quantitative researchers sometimes fail to correctly calculate standard errors and sometimes fail to include them at all (and thus fail to judge the uncertainty of their inferences). However, qualitative researchers include this uncertainty estimate much more rarely.

Indeed, the problem in the traditional presidency literature is a fairly general tendency to avoid uncertainty estimates, and, when they are given, in most cases they are often far too low even to be taken seriously. It takes much more than a cogently argued point to verify an empirical claim about the world. We do not necessarily need quantitative analyses, but whatever analyses we have must be far more systematic.

Perhaps the most important consequence of incorrectly judging or failing to report estimates of uncertainty is the overwhelming urge many presidency scholars feel to make prescriptions—often suggesting fundamental changes in the structure of American democracy, electoral terms, White House organization, presidential policy making, the rules that govern the press, and virtually every other aspect of the subject. Wayne (1983) and others have written about the "almost evangelic tone of much of the literature." In presidency research, we have the luxury (and drudgery) of knowing that many of our recommendations will not be implemented. Nonetheless, prescriptions without adequate judgments of uncertainty are just as irresponsible. If we are listened to at some point, as we occasionally are, improper uncertainty estimates might cause policy makers to act too early, perhaps doing significant damage by creating political instability or even civil war.[2] Prior to making prescriptions, we should be asking ourselves whether we are willing to risk the unintended or unknown consequences of proposed institutional reforms.[3] Neustadt and May (1986, 274) propose a useful method of encouraging policy makers to judge the uncertainty of their conclusions. They ask, "How much of your own money would you wager on it?" The varying utility of money across individuals prevents this from becoming a universal metric for gauging uncertainty, but it does drive home the importance of the judgment.

Several of the first-round authors recognize a consequence of this problem when they complain that the subfields they are reviewing "conflate what is with what ought to be" (Hult, 1990, 52). Rose (1990, 52) even finds some situations in which the two should *not* be distinguished. In my view, presidency scholars should refrain from making prescriptive statements for some time—at least when it comes to most of the critically important problems we study. We should only move to prescription when we have reduced uncertainty far more than the current situation. If we followed *this* prescription, we would not conflate the two at all, but presidency researchers would still have plenty of important subjects to write about and aspects of the presidency to analyze.

I provide an illustration of this point below, albeit with some apprehension, since my purpose is to criticize a research tradition and to suggest possible improvements, not to pick on anyone in particular. In fact, most of the first-round papers meet the highest standards of the traditional presidency literature.

With this focus clearly in mind, consider Paul Light's chapter as an example of the general point. Light makes the argument that presidential policy is increasingly characterized by short-term rather than long-term policy making. Unfortunately, he provides no systematic empirical evidence for this claim but does recognize this in the following qualification:[4]

> Luckily for the poor researcher who would have to operationalize this variable, this distinction between short-term and long-term policy is offered less to create a new empirical category for research and more to illustrate the primary theme of this brief paper. . . . I am convinced that careful coding would find an increase in short-term policy vis-à-vis long-term policy over the past two decades, for presidential policymaking is increasingly characterized by a set of incentives and constraints which reward short-term policy and make long-term policy almost impossible. (Light, 1990, 6–7)

Light spends much of the rest of his paper exploring the *reasons* why presidential policy making is increasingly oriented to the short term, and he concludes with some wide-ranging prescriptions. These include "election reforms, more evaluation capacity, a sensitivity to



the impact of monitoring on implementation," in addition to electing "an occasional prophet" (ibid., 42).

Light is prepared to accept the possibility that, if he is wrong about the prior empirical claim, most of his paper is superfluous. We might legitimately criticize him for climbing so far out on a limb not known to be anchored at all to the tree. Of course, this criticism is certainly correct, but a far more important point is the strategy of inquiry his paper represents. The strategy is one in which the author puts much effort into the interesting questions and does not always take sufficient time to verify the prior empirical claims on which these questions stand. Studying the prior questions would not be as exciting, but they are essential. Unfortunately, this research strategy is inherent in too much of the traditional presidency literature.

One of the key contributions of science is that it helps keep us from tricking ourselves. Human beings are extraordinarily good at pattern recognition, but we are not very good at verifying when patterns we perceive are more apparent than real. Is Light's basic assumption correct? Suppose for a moment that he is wrong, that the proportion of short- and long-term policy has not changed in the last two decades. How could we reconcile this supposition with the fact that we perceive a change toward short-term policy? Perhaps presidency scholars, like everyone else, remember events that happened yesterday more easily than events of twenty years ago. Perhaps we remember presidential policies from 1970 that *persist* to today—by definition, long-term events—and forget some of the short-term policies that have little bearing on today. From Nixon, we remember Vietnam more vividly than his politically expedient opposition to busing in order to placate the South, but from Bush we remember the flag-burning issue as easily as the end of the Cold War and the invasions of Panama and Grenada. Thus, perhaps our casual observation that short-term presidential policy making has increased is only a natural consequence of our necessarily limited memory. A more systematic analysis might indicate that no change has occurred at all.[5]

Light's argument seems plausible enough to me, but the point of science is that we should not trust our casual judgments about such matters. We should do the hard work implied in Light's above quotation. In addition to Light's preliminary analysis, someone should provide a clear definition of short-term and long-term policy and then catalog decisions from presidential diaries or even speeches.

(The process of collecting these data systematically will further define and thus strengthen the theory of short- and long-term policy.) Only with a replicable analysis such as this can we be sure of the basic empirical point. This sort of research might be less fun, and perhaps even less interesting, but its results would be no less important.

This argument is relevant to one of many interesting debates that has been ongoing among scholars in American politics (and was revisited several times during the Presidency Research Conference). The question was why the congressional literature is systematic and theoretically and empirically advanced, whereas the presidency literature seems to lag so far behind. Many reasons are offered for this puzzling difference: more data exist in the study of Congress; members of Congress are more accessible to scholars than is the president; the presidency literature is filled with books written by numerous participant observers without much systematic rigor. In my view, each of these is partially true, but in the end none can account for the difference between the two literatures: much data exist in presidency research too (King and Ragsdale, 1988); the president is not very accessible, but most of his staff are (as Kessel, Pika, Peterson, and others have repeatedly shown), and, regardless, most congressional scholars do not conduct personal interviews with the people they study—the congressional literature also includes participant observer reports, and numerous presidency scholars do research other than participant observation.

In my view, what accounts for the difference between the congressional and presidential literatures is that in the former, but not the latter, scholars spent considerable time recording systematic, but descriptive, patterns. For example, Erikson (1971) and Mayhew (1974) first identified the increasing incumbency advantage and corresponding decline in competitiveness of congressional elections. They did this with fairly systematic and quite careful research, but the literature did not stop there. Literally dozens of scholars published articles showing nothing more than the increase in incumbency advantage. The results were duplicated, replicated, verified, and made much more precise. After scores of descriptive articles, we were quite certain that congressional elections were becoming less competitive and that incumbents were getting a larger share of the vote just because they were incumbents. Only then could serious work begin on building theories to explain these phenomenon, and theory building did begin. The result is that we now have well-developed theories, reasonably strong evidence for them, and a vibrant and active literature on congressional elections.[6] However, I believe the key to the causal explanations we all admire was this prior systematic descriptive work.



I see no reason why we could not follow this same research strategy in the presidency literature. One important possibility is actually Light's suggestive research. For example, suppose that the systematic descriptive work I am advocating were to indicate that Light is correct, that presidents spend considerably more time on short- than on long-term policy making now than twenty years ago. This would be an incredibly important finding. Scholars in our sister subfield spent fifteen years studying a decline in the competitiveness of House elections because this was a key element of democracy. If we could show that Light is correct—that modern presidents spend more time on short- than on long-term policy—we would have a much more important discovery about American politics and American democracy. A whole industry of political scientists might form, trying to explain the newly discovered trend, just as it did in the congress subfield more than a decade ago. However, without this *systematic descriptive* work, Light's (and other's) suggestive but hypothetical speculations will only encourage other speculations and not lead to the kind of general explanation and understanding that we all seek.

## *Toward Incorrect Theories*

I do not disagree with much of the substance in the first-round papers. In part, this is because they are all reviews and, in part, because all do a good job in their stated purpose. However, it is difficult to find much of anything in the literature these papers review that one could disagree with, even in principle. Indeed, from one perspective, the big problem in this literature is its goal: everyone seems to be searching for richer theories, more detailed contextual description, and more all-inclusive theoretical concepts. This goal is precisely what we need for some purposes but exactly the opposite for others. Before explaining these purposes, I provide a few examples.

In describing a "model of political personality," Hargrove (1990, 9) writes, "it must be sufficiently broad to permit its use by advocates of contending theories and those who do not use personality theory at all." He tells us repeatedly to "avoid psychological reductionism" (17) or "try to avoid reductionism" (23). Finally, his summary advice to the subfield is along precisely the same lines: "Avoid reductionism in explanations and look for congruence in motivation. Combine both person and context in explanations. Search for abstraction and generalization but respect the individuality of lives and styles. Compare presidents in ordinary language that will be consistent with diverging theories" (50).

Hult (1990, 12) also argues that, despite the richness of research on presidential advisory systems, we need more richness: "Other efforts to categorize whole advisory systems and to classify presidential management styles fail to 'capture the complexity and variation in advisory practices' (Greenstein, 1988: p. 351)." And at more length and even more explicitly, she writes:

> Scholars have considered numerous independent variables in trying to account for advisory arrangements and the impact of advice on presidential actions. These dimensions have ranged from presidential styles, ideology, and strategy to small group interactions and staff structures to prevailing values in the larger political system, fragmentation in Washington, and features of the relevant policy arena. Despite the richness of presidential, organizational, and environmental factors, much work concentrates on only one cluster of variables (Hult, 1990, 46).

In one way or another, every one of the first-round authors (and many other authors in the rest of the presidency literature) is arguing for additional richness of theory, explanation, and context.[7]

In judging this methodological advice, we must distinguish *social science* from *history and biography*. The difference I have in mind is not entirely disciplinary, since many historians make important contributions to social science and numerous social scientists describe themselves as primarily historians. Rather, I am interested in the specific purposes of our scholarship, and from this perspective, most of the traditional presidency literature is composed of first-class historical and biographical accounts of presidents and their administrations. The dominant goals include getting the facts right, chronicling the fascinating stories in and around the White House, and comparing and contrasting different presidents, their aids, policies, successes,



and failures. In pursuing these largely descriptive goals, no field in political science or American politics is more developed. Few across the social sciences match the extent of our descriptions. The history of no other institution or person is as completely recorded as that of the American presidency and president. For this, presidency scholars should be justifiably proud. The frequently expressed pessimism scholars have about the presidency field cannot be about its main goal of history and description.

On the other hand, presidency research has a way to go in pursuing the goals of social science. As evidence of how far other fields have surpassed us on this score, consider that traditional presidency scholarship rarely even makes it into mainstream political science journals anymore. A parsimonious explanation, and not richer and richer contextual description, is the immediate goal of social science. Rich description is important for understanding what is to be explained by later systematic analysis; it is important for telling history, where we would prefer to know the facts as closely and perhaps even as completely as possible. But it is not particularly useful as theory or explanation.

To make the distinction between social science and history clearer in this context, let us analyze the scholarly benefit we would get from collecting a large number of explanatory variables. For historical purposes, this is a very useful activity, since it will help a scholar to provide a more accurate picture of a president or the presidency. However, *for social science, no reasonable argument can be constructed in which all conceivable explanatory variables could be used at once in making inferences*. In some specific cases, other explanatory variables will help. In general, however, they will not. (See King, 1991).

Take, for example, a very good and specific testable hypothesis given by Hult (1990, 9): "Staff units that span White House boundaries (e.g., press, congressional liaison) will be less influential in White House discussions, since the president and other staffers frequently view the contacts with outsiders with some suspicion. Whether this is so needs to be tested more systematically." The dependent variable in Hult's hypothesis is staff unit influence in White House discussions, by which she presumably means the influence the head of a staff unit has on the heads of other staff units and the president. It would take some work to measure this, but it seems feasible to measure it—perhaps with Kessel's (1984) interview and network analysis data or some other type of information. The unit of analysis is the staff unit, and the key explanatory variable is whether or not a staff unit spans White House boundaries.

How would it help us in assessing the effect of spanning boundaries on influence in the White House to simultaneously study the effects of other explanatory variables? For the answer to this question, I turn to the quantitative literature, where the problem is known as omitted variable bias. My point is not to turn this into a quantitative inference in this qualitative research project. Thus, for simplicity, I consider two other possible explanatory variables: (1) number of years in which the organization of a staff unit has remained the same, and (2) the quality and experience of the head of the staff unit.

It might be interesting to know the effect of the number of years since reorganization. Indeed, even listing the relative durations between reorganizations for each staff unit could provide a gold mine of descriptive information. It would even be easy to imagine a study trying to describe or explain these interagency differences. Moreover, this variable would seem to have an important effect on the dependent variable, because stable staff unit organizations will probably have deeper institutional relationships with other staff units, and influence will be easier and more frequent. Staff units that have recently had institutional overhauls will likely be more preoccupied with internal politics than with influencing other staff units.

However, it turns out that this variable is largely worthless in gauging the effect of spanning boundaries on staff unit influence, and it will therefore have no impact on tests of this hypothesis. The reason is that organizational stability is unrelated to the key explanatory variable, spanning boundaries, unless a particular president specifically set out to reorganize staff units that spanned boundaries more (or less) frequently than other staff units. Since this is not usually the case, any estimate (quantitative or qualitative) of the effect of spanning boundaries on a staff unit's influence in the White House will remain unbiased regardless of the number of years in which the staff unit's organization has remained the same. Gathering information on this variable might be interesting by itself, but it would be a waste of resources for purposes of examining Hult's hypothesis. Thus, however interesting this variable might be for descriptive purposes, we



can and probably should omit it from our explanatory analysis, and our research will suffer no negative consequences.

In testing the effect of spanning boundaries on White House influence, we would probably not simply compare the influence of those staff units that spanned boundaries with those that did not. This would not be advisable, because the heads of these staff units are likely selected in ways that are related to their probable influence. For example, suppose that Hult is correct in postulating that staff units that span White House boundaries will be less influential. A logical consequence (or a rational expectation, in economics terms) would mean that the best staff members would be appointed to staff units that do not span boundaries. If one just compared the influence of the two types of staff units, we might find that those that spanned boundaries are less influential, but this might be due solely or primarily to the initial staff appointment process. Indeed, the better the initial appointment process was in incorporating these rational expectations, the more biased would be the estimate produced by this (naive) study.[8] One way to cope with this possible confounding problem is to measure the quality or experience of the head of each staff unit.

In a quantitative analysis, we might well include this second variable in a regression equation to avoid omitted variable bias.[9] However, the same bias would afflict a qualitative analysis if one did not somehow take into account the quality of the staff head. If one possessed some assessment of how good each staff head was—and we would have much better information on this variable in a qualitative rather than a quantitative study—one could choose cases where the head was very experienced or skilled for both a staff unit that spanned boundaries and one that did not. We might also choose both types of staff units with inexperienced or less skilled heads. If Hult is correct, we would find that the spanning boundaries variable still explained influence in the White House, even after we held constant the effect of quality and experience of the personnel.

Thus, collecting some types of additional explanatory variables will help us explore the implications and veracity of our key hypotheses. However, a wide-ranging collection of all possible explanatory variables will not help us achieve any relevant goal of social scientific causal inference. In fact, it is even worse, since each additional variable for which we simultaneously estimate a causal effect reduces the precision of all of our causal estimates. Thus, "success" in amassing a larger and larger number of variables will automatically produce failure in learning about any one causal inference. This strategy would still be interesting and productive for descriptive or historical purposes, but it would not be very useful for social scientific inference.

I underscore what we all probably know here, that good history is not necessarily good social science (and vice versa). A theory that is sufficiently broad, or an "explanation" with a huge number of explanatory variables, illuminates nothing whatsoever. A "theory" incapable of being wrong is neither a theory nor an explanation. Philosophers of science might say that a theory or hypothesis should be "falsifiable," but this concept confounds two separate criteria. The first is that, in principle, any theoretical statement should be either true or false; in particular, a theory that purports to combine every perspective about a subject is unlikely to be capable of being false. In other words, if one cannot imagine a series of events in the real world that would convince one that a theory is wrong, then the theory is useless from the start. Second, research methods capable of distinguishing whether a theory is true or false must be identified.

In my view, the first criterion is essential. The second should be used to judge the veracity of particular theories, rather than these theories per se. Separating the two criteria is pivotal. Traditional presidency researchers are sometimes good at stating theories capable of being false; they have rarely provided sufficient evidence to support or oppose any particular theory. Completing the first step alone is useful, provided it is recognized as only the first step—and as long as one reports it with a fairly large estimate of uncertainty.

In order to state theories in such a way that they are capable of being false, we must choose theoretical statements that are *specific*. For example, consider the following quotation from Hargrove's paper (1990, 6), outlining the purpose of his review of the presidential psychology literature. I emphasize those words or phrases in this passage for which the presidency literature provides an insufficiently precise means of understanding, much less measuring with any degree of specificity:[10]

> This essay will explore two areas in which the consequences of personality are very important:
>
> 1. Presidents who find themselves in situations which place *high stress* on them may *deal with the stress* by *responding to internal psychological needs* rather than *acting in ways appropriate to the external situation*.
>
> There are difficult interpretive questions here because there is usually more than one appropriate response to a situation. *Ego defensive actions* that serve the *vulnerable person* may be identified in an individual if they occur often enough and are accompanied by *recognizable emotions* that *reveal the vulnerability*.
>
> 2. The *management of decision-making* among lieutenants and advisers is a manifestation of *style of authority* which is, in part, a reflection of *self-confidence* and *cognitive openness* as well as *beliefs about effective leadership*.



The problem with the emphasized words and phrases is not that quantitative measures have not been developed. For most of these concepts, this may never occur. Certainly the rules that do exist for defining these concepts are not adequately precise; a scholar from outside the presidency literature could not come along and identify "cognitive openness," "ego defensive actions," and the other terms, just from the definitions given and perhaps not even from a complete review of the presidential psychobiography literature. We must have rules so that other researchers can come along and apply the same rules and reach the same judgments.[11] Better theory will also come from more precisely laid-out rules.

One possible objection to my argument here is that these concepts, however fuzzy, are what interest presidency researchers. My response is that this could not possibly be true. What interests presidency researchers, by definition, are presidents and the presidency. How these concepts relate to presidents or the presidency is an argument for their proponents to make. Sometimes these arguments have been successful and sometimes not (as Hargrove makes clear in the rest of his paper). However, if the argument cannot be made in particular cases, of what use are fuzzy concepts? How can an ill-defined problem even be interesting in the first place without some specificity? In general, it cannot. However, in much of this literature, specific ideas do underlie much of the argument. The problem is that different scholars have used different psychological terms in the same way or the same psychological terms in different ways. This strategy has often been useful for description—perhaps its main purpose—but it has not always been helpful in arriving at reliable causal inferences.

## *The President as the Unit of Analysis*

Perhaps the best-known methodological problem in the presidency literature is the $n = 1$ problem. This is the idea that only one president is in office at any one time, and so inference is inherently difficult if not impossible. How are we to use John Stuart Mill's methods of difference or agreement if we cannot find two presidents who are alike in all respects but our key explanatory variable? Too much of the world changes when the president changes. Franklin Roosevelt was something like John Kennedy, but how can we hold "things" constant in studying any explanatory variable when "things" include differences in age, health, a world war, the economy, and a myriad other factors. The $n = 1$ problem guarantees that this sort of difficulty will always come up in presidency research, perhaps even more frequently in this subfield than anywhere else.

The basis of the $n = 1$ problem is precisely correct, and it does have exactly these consequences. However, not only is the problem not unique to the presidency literature, but it is perfectly general. It is merely a clear statement of the very definition of causality. To be a little more general about this, Holland (1986) describes what presidency scholars call the $n = 1$ problem as "the fundamental problem of causal inference." To best describe this problem I provide a *definition* of a causal effect, entirely independent of the difficulties we might have in *estimating* this effect.

For simplicity, consider the causal hypothesis that presidents who were once members of Congress veto legislation less frequently. The precise definition of this casual effect is as follows. Consider one president, say Jimmy Carter, and observe the number of veto orders he signs during his four years in office. Then take the same president, turn the clock and the world back to 1976, alter Jimmy Carter so that he is alike in all respects except that he served as a member of Congress, make everyone (but you!) forget the first experiment, and run the world a second time. The causal effect is then the difference between the numbers of vetoes from these two experiments. It should be obvious that one cannot *know* the causal effect even in theory, since one of the values required (the number of vetoes for each experiment) is always unobserved. Thus, this is indeed a fundamental problem, and it should be clear that it is completely general—applying to all types of causal effects. Getting around this fundamental problem is difficult, perhaps even more difficult in the presidency literature, but the problem is not unique here.



In the remainder of this section, I show that the common practice of using the president as the unit of analysis for causal inferences is extremely unlikely to yield reliable empirical conclusions. I then provide some examples of research designs we can use to get around these problems. The president is used as the unit of analysis in studies of decision making, organizational style, and advice structures, but perhaps its most common application is in presidential psychology. Barber's (1980) famous work uses presidents as the unit of analysis; indeed, most analyses that put presidents in categories do the same. Probably the most important findings of the presidential psychology literature are based on this kind of research design. Such analyses predict that presidents with a particular personality profile (referred to in various forms as active-negative, ego-defensive, etcetera) will "rigidify" sometime during their terms in office. Woodrow Wilson and Richard Nixon are generally the two leading examples of this personality profile, and the League of Nations debate and Watergate, respectively, are generally provided as evidence of rigidification (see George and George, 1956).

Many object to these sometimes vague theoretical constructs in the political psychology literature.[12] However, in the analysis that follows, I make the very conservative and optimistic assumption that we have worked out all of the definitional problems associated with ideas like *rigidify* and *ego-defensiveness*. Nevertheless, even in this rosy situation, taking the president as the appropriate unit of analysis is still fraught with problems.

Let us take a very simple case. Suppose we have a single dichotomous explanatory variable. To fix ideas, let this variable be a presidential personality type, either "good" or "bad." Let the dependent variable be whether or not a president rigidifies during his or her term. The basic idea is that the probability of rigidifying is larger for "bad personality" presidents than "good personality" presidents, although I emphasize again that this logic applies for any research problem in which the president is the unit of analysis. In particular, the hypothesis is that the following difference in probabilities is positive and large enough to make a substantive difference:

$$\Pr(\text{rigidify}|\text{bad}) - \Pr(\text{rigidify}|\text{good}).$$

I refer to this difference in probabilities as the *causal effect* we are interested in. Note that it is also another example of the fundamental problem of causal inference, since we can never assess both probabilities for a single president. Furthermore, although we might estimate this effect with either quantitative or qualitative methods, the underlying logic of inference about this effect is essentially the same when you use the president as the unit of analysis: One does a number of case studies of presidents and calculates the proportion of bad presidents who rigidify and subtracts that from the proportion of good presidents who rigidify. This *estimated effect* is based on however many presidents are included in our study. Whether the study is quantitative, and measures are numbers, or qualitative, and measures are just verbal evaluations, this same inference is the goal of the analysis.[13]

My question is how good the estimated effect can possibly be when it is based on the small number of presidents we have available. To answer this question, we need to calculate the "statistical power" of this estimator. In particular, for a given effect size, standard methods of inference enable us to calculate the number of observations (presidents) necessary to find a .05 significant difference some fixed proportion of the time. This fixed proportion is referred to as the power, and the 0.05 significance level is just a reasonable (and arbitrary) convention in most of the social sciences. Table 12 reports some of these calculations for effect sizes of 0.1 and 0.2—reasonably large effects for political science and enormous ones for a field as necessarily imprecise and uncertain as the psychobiography of presidents (after all, presidency scholars rarely even get to interview presidents, much less psychoanalyze them in repeated individual meetings). For each of these effect sizes, the table reports the number of presidents we would need in order to find a significant difference at the 0.05 level 80 percent, 90 percent, and 95 percent of the time, respectively.

The results in table 12 should be somewhat disconcerting to anyone using the president as the unit of analysis to test any hypothesis. Even if we analyzed every existing president, we would still not come near the required number of observations necessary to make reliable inferences with a reasonable degree of certainty. In practice, of course, it takes considerable documentation to conduct a reliable psychological analysis of a president, and most researchers therefore use only postwar presidents. The table also shows that the lower the statistical power one is willing to live with, the fewer the presidents that would be necessary to find a significant difference at the .05 level. Also, the larger the true effect, the fewer presidents there would have to be; this is a more difficult criterion in practice, of course, since the purpose of the empirical analysis is to estimate the effect, so it cannot be known prior to that analysis.



TABLE 12

*Required Sample Sizes to Use the President as the Unit of Analysis*

| | Difference in Proportions | |
|---|---|---|
| Power | .1 | .2 |
| .80 | 388 | 91 |
| .90 | 515 | 121 |
| .95 | 638 | 150 |

To make this difficult situation somewhat more graphic, suppose we waited until enough presidents had served to enable us to use the president as the unit of analysis. Under the most optimistic of research circumstances, let us suppose that we had only one-term presidents from now on and we were somehow able to analyze every president beginning with George Washington. We would also have to assume that the basic structure and validity of the hypothesis, as well as the size of the causal effect, remains constant, and that no explanatory variables omitted from this analysis would cause any bias. If any of these rosy assumptions are incorrect (and most certainly are), then the conclusions of this section will be even more pessimistic regarding any efforts to use the president as the unit of analysis. Under these assumption, table 13 reports the approximate *year* in which a reliable analysis could be conducted, in that a significant effect would be found if it indeed did exist.

Waiting until somewhere from the year 2193 to the year 4378 is obviously not feasible for a dissertation project or any other scholarly endeavor! If the true effect were much larger than .2, we might be able to get away with a much smaller number of presidents in the analysis, but we should not choose a research design that is only

TABLE 13
*Year in Which the President Can Be Used as the Unit of Analysis*

| | Difference in Proportions | |
|---|---|---|
| Power | .1 | .2 |
| .80 | 3378 | 2193 |
| .90 | 3888 | 2313 |
| .95 | 4378 | 2428 |

useful if the effect we are estimating is enormous or that requires us to wait until nearly the end of time.

In his first-round paper, Hargrove writes, "One cannot have it both ways, arguing that individuals are not important and then in the same breath arguing that the importance of individuals precludes systematic study" (1990, 2). This point about the critics of the presidential psychology is right on the mark. Indeed, it is almost certainly true that individuals are important *and* that presidents can be studied systematically. However, it is clear that the systematic study of individual presidents should not continue in the tradition of using the president as the unit of analysis. If it is advisable to give up this research design, one need not necessarily give up political psychology or the study of the other areas that rely on the president as the unit of analysis.

The solution is to look for ways of multiplying the number of observations—looking for additional observable implications of the same theory. For example, an obvious choice for political psychology and for decision-making research is to use the decision as the unit of analysis. One possibility is to use the president's decision about whether to endorse each piece of congressional legislation or to veto it. This is a set of easily identifiable decisions and might enable one to generate a larger number of specific predictions from a psychological perspective. Even twenty-five or fifty observations would provide a significantly improved research design. If this study goes well, one might be able to abstract the essential features of the psychological variables, making it easier to collect an even larger number of observations. Any prediction along these lines would obviously not be correct in every instance, and the more observations we collect and the less time we have to devote to any one, the more error there will



likely be. However, all we should expect is that the difference in proportions (the estimated effect) be large enough to be important and detectable.[14]

Can the study of presidential personalities or decision-making styles help us to predict decisions such as these? Since the president is not offered exactly the same choice in each instance, the degree to which each situation would generate ego-defensive behavior would not be identical even for the same president.[15] The very process of deriving coding rules would help make the theory more specific and useful, and the study itself would help us to know whether we are indeed right. If political psychology turns out not to be able to make predictions in an area like this, where the decision context and possible outcomes are very clear, then perhaps one might conclude that we should channel our efforts in more rewarding directions. Of course, on the basis of the extensive literature in this area, I feel fairly certain that this study would yield fairly strong effects, and if so, it would give presidential psychology studies considerably more specificity and thus significantly more reliability.

A similar problem to that in the presidential psychology literature is in the literature on leadership, as Sinclair ably argues in this volume (chap. 5). She asks the basic question of whether presidents lead. Restated in causal language, we might say, if another person were in the same situation, would he or she decide something different? Because of the fundamental problem of causal inference, this would obviously be a particularly difficult problem to study in presidency research. It would be very difficult to get around this problem even if we used decision points, instead of presidents, as the unit of analysis.

Perhaps, then, *we can look for a different set of observable consequences of the same theory*. For example, we might construct the following experiment. We could go into the archives of some president and extract a number of cases where there were lots of evidence about advice from presidential staff. We could then change a few key facts in this situation so respondents would not remember what happened in the actual historical event. Alternatively, and probably better, we could find some ordinary presidential decisions. This would be better because the experimental subjects would be less likely to remember and especially because more presidential decisions are indeed fairly ordinary. We could then give the set of advice from the various advisers to a few dozen people—perhaps randomly selected citizens, or perhaps people who seek to understand the role of the president (such as graduate students or law students), or those who were close to presidents or would like to be president (such as former presidential staffers or members of Congress). There are many questions we could then ask. First, if there was very little variation in the decisions made by these people, then the evidence would be consistent with the idea that presidential leadership does not exist: anyone would make the same decision in the same situation. If the more likely situation occurred, where there is some variation, could we account for it? Is the variation due to the person's party affiliation, education, occupation, degree of ego-defensiveness, or something else? With a study like this we could generate a large number of observations and would likely learn many important facts about leadership, decision making, and even presidential personality.

One final way to increase the number of observations is to look for observable implications of one's theory at other levels of aggregation. For example, one of Aldrich's questions is the extent to which the presidential campaign matters. Our forecasting models seem to do quite well without any information from the campaign (see Rosenstone, 1983). Indeed, contrary to media expectations, Dukakis did better than our prediction models indicated in 1988. Because of the Fundamental Problem, in order to estimate the causal effect we obviously cannot measure the observed election result in 1988 *and* in a hypothetical election that was the same as 1988 except for certain key campaign events.

However, campaign events such as the Willie Horton ads presumably had different effects in different states. If the ads were successful at provoking latent racism among whites, then the electoral effect would presumably be different in Washington, D.C., with a very high proportion of blacks, than in New Hampshire, with almost none. Remember that we really do not care whether this effect is different in different states per se, but since this idea is consistent with the more general hypothesis, this study could provide the necessary critical evidence. This would not even necessarily be a full quantitative study, since we could learn an awful lot from a map with prediction errors in various states highlighted; one could then use our enormous base of qualitative knowledge about the campaign in different states to try to make sense of the results.



We could even go farther and look at individual level survey data. This might help us determine whether racist whites voted for Dukakis less frequently than other whites in similar economic and political circumstances. Again, if we are mainly interested in the effect of the campaign, a survey is not directly relevant, but its indirect relevance in possibly confirming an important implication of the more general theory is overwhelming.

Although one cannot use the president as the unit of analysis in order to derive reliable causal inferences, most of the theories that have been explored in this way could be studied in other ways too. Perhaps the most productive strategy is to search for numerous additional observable implications of one's theory. These implications may be at different levels of aggregation, in different political systems, at different times, or with different measures of the same variables.

## *Concluding Remarks*

The traditional presidency literature has accomplished an enormous amount in the area of history and contextual description. However, progress in a social science of the American presidency is far less advanced.

Future research needs to emphasize not only quantitative analyses but rigorous and systematic qualitative research. We need to insist absolutely that any prediction or explanation must come with a fair assessment of its uncertainty. My example of research based on the president as the unit of analysis demonstrates just how uncertain some of our best work is likely to be even in the foreseeable future. Continuing research will undoubtedly reduce this uncertainty, but only if our qualitative research follows the standard rules of scientific inference.

## *Notes*

1. This is a specific version of the argument in King, Verba, and Keohane (in progress).

2. Without more systematic research, there will be many situations where the world would be better off without any presidency literature. Florence Nightingale once said at a minimum, hospitals should not *spread* infection."

3. Perhaps it pays to remember that the goal of the academic members of the McGovern-Fraiser Commission was to increase party discipline and organizational strength—precisely the opposite of what happened.

4. This discussion characterizes the first version of Light's paper. He has subsequently revised his paper to include some quite suggestive empirical evidence, though still not conclusive even by his judgment.

5. Indeed, what president in modern times was more concerned with short-term political gain than Nixon? Watergate could easily be interpreted as a series of presidential decisions seemingly designed for no purpose other than short-term political gain.

6. The newest work in the congressional elections literature displays a similar trend designed to describe and explain the apparent increase in divided government at the federal and state levels.

7. Even Aldrich is arguing for more complete, realistic, and all-encompassing theories (1990, 48): "The division into who runs, who is nominated, who is elected, and how does all this affect the actions of the victor is, unfortunately, too tidy. . . . A theory of one component is necessarily incomplete. The theory of presidential selection must attempt to answer all four of these questions." Unlike the other presidency subfields, which I have been referring to as the traditional presidency literature, Aldrich's goal may actually be more immediately reachable, because scholars have been very specific along the way and have made considerable progress on separate theories and empirical analyses from each of the four parts.

8. This is the same kind of selection bias that occurs when one tries to predict success in graduate school on the basis of standardized test scores or undergraduate grades, using as data only those students who were admitted. It turns out that the better the admissions committee was in choosing the best students, the poorer job we would think they did.

9. More specifically, one can avoid omitted variable bias if the omitted variable is uncorrelated with the included one or the omitted variable has no effect on the dependent variable. If neither condition holds, then omitting this variable will bias one's estimate of the included variable's effect on the dependent variable.

10. Hargrove is quite aware of the problems with defining these concepts, as he makes clear from the first sentence of the third paragraph of this quotation. I certainly do not mean to blame him for problems in the political psychology literature!

11. Social scientists should not measure concepts the same way as the Supreme Court determines whether something is pornographic.

12. Did Nixon and Woodrow Wilson really rigidify? Or did they think that they could really get what they wanted? Many argued that the game was over, and yet these two pursued their goals (staying if office and in the League of Nations, respectively) relentlessly. Perhaps this is true, but if these goals were of incredible importance to them, perhaps it made perfect sense for them to keep pushing. From this perspective, Nixon was fighting to save face, to survive politically, and to keep a positive place in history. One would not have to stretch the bounds of rational choice modeling very far to fit in the behaviors of these two political actors. Wouldn't you fight as hard as Nixon? Perhaps he did not rigidify but instead became more and more flexible and creative about trying to get around a very big roadblock. If he managed to stall until the end of his term, would we be calling this *rigidity* or *cleverness?* If we cannot agree on even vague measures of either the explanatory or the dependent variables, how can we expect to find a relationship among them that anyone would accept?



13. This procedure is valid so long as we are not omitting key explanatory variables that are prior to personality and correlated with it and also that affect the probability of rigidification in office. Since the theories of personality usually used in this literature assume that personality is formed at an extremely early age, the study would seem valid in design.

14. Another possibility is to use decisions to appoint and dismiss various White House personnel.

15. Bensel (1980) did an analogous study where he coded each piece of congressional legislation to see if it narrowed the discretion of bureaucrats.

## *References*


Aldrich, John. 1990. "Presidential Selection," paper prepared for the Presidency Research Conference, University of Pittsburgh, 12–14 November.

Barber, James David. 1980. *The Presidential Character.* N.J.: Prentice Hall.

Bensel, Richard. 1980. "Creating the Statutory State: The Implications of a Rule of Law Standard in American Politics." *American Political Science Review* 74(3):734–44.

Edwards, George C., III. 1983. "Quantitative Analysis." In *Studying the Presidency,* ed. George C. Edwards, III and Stephen J. Wayne. Knoxville: University of Tennessee Press.

Erikson, Robert S. 1971. "The Advantage of Incumbency in Congressional Elections." *Polity* 3:395–405.

George, Alexander L., and Juliette L. George. 1956. *Woodrow Wilson and Colonel House, A Personality Study.* New York: John Day.

Greenstein, Fred I. 1988. *Leadership in the Modern Presidency.* Cambridge: Harvard University Press.

Hargrove, Erwin C. 1990. "Presidential Personality and Political Style." Paper prepared for the Presidency Research Conference, University of Pittsburgh, 12–14 November.

Holland, Paul, 1986. "Statistics and Causal Inference." *Journal of the American Statistical Association* 81:945–60.

Hult, Karen M. 1990. "Advising the President." Paper prepared for the Presidency Research Conference, University of Pittsburgh, 12–14 November.

Kessel, John. 1984. "The Structure of the Reagan White House." *American Journal of Political Science* 27:431–63.

King, Gary. 1991. " 'Truth' is Stranger Than Prediction, More Questionable Than Causal Inference," *American Journal of Political Science* 35 (November): 1047–53.

King, Gary, and Lyn Ragsdale, 1988. *The Elusive Executive: Discovering Statistical Patterns in the Presidency.* Washington, D.C.: Congressional Quarterly Press.

King, Gary, Sidney Verba, and Robert O. Keohane. In progress. *Scientific Inference in Qualitative Research.*

Light, Paul C. 1990. "Presidential Policymaking." Paper prepared for the Presidency Research Conference, University of Pittsburgh, 12–14 November.

Mayhew, David R. 1974. "Congressional Elections: The Case of the Vanishing Marginals." *Polity* 6:295–317.

Neustadt, Richard E., and Earnest R. May. 1986. *Thinking in Time: The Uses of History for Decision-Makers.* New York: Free Press.

Rose, Richard, 1990. "Evaluations." Paper prepared for the Presidency Research Conference, University of Pittsburgh, 12–14 November.

Rosenstone, Steven R. 1983. *Forecasting Presidential Elections.* New Haven: Yale University Press.

Wayne, Stephen J. 1983. "An Introduction to Research on the Presidency," in *Studying the Presidency,* ed. George C. Edwards III and Stephen J. Wayne. Knoxville: University of Tennessee Press.