Political Psychology or Politicized Psychology: Is the Road to Scientific Hell Paved with Good Moral Intentions?

Philip E. Tetlock  
University of California at Berkeley

This article proceeds from the premise that a completely value-neutral political psychology is impossible. Testing hypotheses about the efficacy of deterrence or the pervasiveness of racism or the quality of decision making inevitably requires value-charged trade-offs between Type I errors (rejecting the null hypothesis when it is true) and Type II errors (failing to reject the null hypothesis when it is false). The article goes on, however, to argue that our collective credibility as a science depends on self-critical efforts to monitor and minimize the influence of scientifically irrelevant values on inquiry. I identify two examples of research programs—White’s work on deterrence and the Sears and Kinder work on symbolic racism—in which the moral-political values of the investigators appear to have profoundly shaped standards of evidence and proof in testing competing hypotheses. I also identify logical and empirical strategies that investigators can use to check the influence of extraneous values. These strategies include rigorous skepticism toward counterfactuals that underlie causal claims in historical analyses, embedding of experimental manipulations in representative sample surveys to isolate determinants of public opinion, developing methods to translate case studies into standardized data languages so that we can more readily identify potential sources of bias, and continual open-mindedness to the possibility that patterns of thinking that scholarly observers laud as cognitively or morally superior in one set of political settings may look quite maladaptive or immoral in other political settings. The article closes with a transparently value-laden appeal to preserve the autonomy of political psychology as a science by distinguishing sharply between when we speak for a scientific discipline and when we speak as concerned citizens.

The short answer to the question posed in the title is “yes, good moral intentions can lead us to scientific hell, but there is nothing inevitable or predeter-
minged about that outcome.” It is possible for a self-correcting scientific community to do reasonably rigorous, albeit never perfectly value-neutral political psychology—a claim that I will try to illustrate with examples from a variety of research domains.

Let’s start, however, in the old-fashioned positivist way by providing a key operational definition: What exactly is scientific hell? I use the concept to denote the complete collapse of our credibility as a science. We find ourselves in scientific hell when we discover that our powers of persuasion are limited to those who were already predisposed to agree with us (or when our claims to expertise are granted only by people who share our moral-political outlook). Thoughtful outsiders cease to look upon us as scientists and see us rather as political partisans of one stripe or another.

How do we fall into scientific hell? The principal temptation in political psychology—the forbidden fruit—is to permit our political passions to trump normal scientific standards of evidence and proof. Researchers sometimes feel so passionately about a cause—about the need to avoid nuclear war or war in general or the need to eliminate racism in both its subtle and overt forms—that those passions influence key methodological and conceptual decisions in research programs. When journal reviewers, editors, and funding agencies feel the same way about the cause, they are less likely to detect and correct potential logical or methodological bias. As a result, political psychology becomes politicized.

In this article, I focus on two specific examples of how political psychology can be deeply politicized: Ralph White’s work on the causes of war and peace and David Sears and Donald Kinder’s research program on symbolic racism. I then identify four strategies that investigators can use to minimize (although never completely eliminate) the influence of scientifically irrelevant values on research conclusions. I close with a plea to preserve the “epistemic autonomy” of political psychology as a science.

CLAIMING CREDIT FOR HISTORICAL OUTCOMES: THE CASE OF U.S.-SOVIET RELATIONS

In a series of articles in the 1970s and 1980s, Ralph White made a series of controversial causal claims (see especially White, 1984). He maintained, for instance, that the easing of tensions in the U.S.-Soviet relationship in 1963—an easing that led to the signing of the partial nuclear test ban treaty—constituted evidence for the effectiveness of Charles Osgood’s GRIT strategy of conflict resolution (graduated and reciprocated initiatives in tension reduction). White, moreover, was not alone in making this claim: Amitai Etzioni (1967) had already advanced a similar argument.

A “natural experiment” of this form is riddled with threats to its internal validity. We cannot be sure whether Khrushchev was responding to conciliatory gestures and rhetoric by Kennedy, or whether he was responding to the apparent willingness of the United States to use its massive nuclear superiority to impose its will during the Cuban Missile Crisis a few months earlier, or whether he was responding to domestic political imperatives, or whether Khrushchev was, once again, being quirky, impulsive and idiosyncratic, and the movement toward good relations was just one more blip in the random walk of history. Although White was willing to draw very strong inferences about causality in the last year of the Kennedy administration, he was deeply skeptical of efforts by the Reagan administration to claim credit for Gorbachev’s ascension to power, for democratizing the Soviet Union, and for winning the Cold War. White gives us no defensible theoretical criteria, however, for assessing what constitutes a good historical counterfactual. The question remains: Why is the counterfactual “without conciliatory Kennedy behavior in 1963, no easing of tensions” more inherently plausible or compelling than the counterfactual “without the Reagan buildup of the early 1980s, no internally initiated reform within the Soviet Union”? Indeed, why should counterfactuals count as evidence at all? White’s research program enjoyed considerable influence among peace activists in the 1980s, but it fails to pass basic logical and methodological requirements for “good science.”

In his 1984 book Fearful Warriors, White also argued that the scientific literature lent support to the call for the West to adopt a posture of “minimal deterrence” with respect to the Soviet Union. He maintained “that perhaps 10 fairly invulnerable submarines almost certainly able to destroy the 200 largest cities in the U.S.S.R., would be sufficient to keep the other side from starting a nuclear war. If so, we don’t necessarily need anything else for that purpose—not ICBMs, not Pershing IIIs, not land-based cruise missiles, perhaps not bombers, and not even a number of nuclear armed submarines comparable to theirs.” My intent in quoting this passage is not to quarrel with White’s recommendations (interestingly, they bear a striking resemblance to current Russian thinking on security issues). My point is simply that White ceased at that point to be a political psychologist in any scientific sense of the term. The argument he advanced rested on a complex mixture of moral, political, and strategic assumptions, including the belief that sudden technological advances would not render submarines vulnerable to attack, the belief that Western retaliation via relatively inaccurate submarine-launched ballistic missiles (most of which can destroy cities but not hardened military targets) was a credible threat that would deter attacks, the belief that the principal adversary of the day (the Soviet Union of the early 1980s) was a defensive, risk-averse power which would not exploit this dramatic American move in the direction of unilateral disarmament, and the belief that American ICBMs exacerbate the adversary’s fears of a first strike and
could trigger a preemptive nuclear attack. Political psychology had become policy advocacy.

WHAT CONSTITUTES EVIDENCE OF RACISM?

Turning to work on symbolic racism, we see a research program with the objective of identifying the underlying psychological sources of public resistance to policies designed to promote racial equality. This research program has been built around a fundamental idea: old-fashioned racism has disappeared, but it has been replaced by a new more subtle form of racism—symbolic racism. This new racism, yoking together prejudice and traditional American values, both veils and legitimizes white racism (cf. Kinder & Sears, 1981; McConahay & Hough, 1976; Sears, 1988).

The guiding idea here is a provocative one, also intuitively convincing. Who would deny that people can disguise racial prejudice? Or that they can express it covertly, through support for traditional values? The central difficulty arises in the strategy that symbolic racism researchers have often used to measure the central construct of interest. Because Sears and Kinder have defined symbolic racism as a fusion of antiblack affect and either traditional or conservative values, they have often relied on survey questions that deliberately confound or mix these two elements. For example, in the Sears and Citrin (1982) study of the California tax revolt, survey respondents qualified as symbolic racists if they passed two tests: (a) they agreed that government should not make any special effort to help blacks and other racial minorities because those groups should help themselves; and (b) if they were opposed to the use of forced busing to produce integration in the schools (a two-item scale in which the items correlated less than .20). In another study by Kinder and Sears (1981), symbolic racism was operationally defined, in part, as opposition to affirmative action and, in part, as opposition to busing.

The core psychological premise of symbolic racism theory—that traditional values and anti-black affect fuel opposition to governmental efforts to promote racial equality—may or may not be correct. But the research strategy adopted here labels people who object to busing or affirmative action on purely ideological (race-neutral) grounds as racist by definition. By confounding ideology and racism at the item level (so that conservatives who oppose government regulations and subsidies in general arc ipso facto racist), symbolic racism researchers make it impossible to assess the independent and interactive effects of conservative ideology and anti-black sentiment on the policy stands that the public takes. The typical symbolic racism scale measures a murky mixture of anti-black affect and stereotypes, conservative beliefs about human nature, and conservative values concerning the importance of self-reliance and minimizing dependency on government handouts. A more value-neutral and informative way of proceeding would be to try to disentangle the distinctive sources of resistance to policies to assist various minorities and to set up one's empirical procedures so that conservatives are not automatically guilty of racism.

The symbolic racism research program politicizes political psychology. To label someone a racist, or to label support for a viewpoint as racist, is to pass moral-political judgment on that individual or viewpoint. Standards of evidence need to be exceptionally clear in passing such judgments.

It is instructive in this connection to consider the parable of the symbolic Marxism scale that Paul Sniderman and I introduced in a 1986 article. Symbolic racism scales are intended to identify people who both dislike blacks and are conservative. One can imagine, correspondingly, a symbolic Marxism scale to identify people who dislike businessmen and are liberal. We asked, in our original paper, how the research community would react to conservative researchers who operationalized their concept of symbolic Marxism with items that focused on support for the civil liberties of American communists and on opposition to aid to right-wing governments. I would expect—indeed, I would hope—for a strong negative reaction. Calling people symbolic Marxists on the basis of such evidence would be a clear-cut example of what C. Wright Mills (1940) called social scientific motive-mongering—in this case, the use of political psychological methods to cast aspersions on political viewpoints with which the researchers disagreed.

VALUE NEUTRALITY: IMPOSSIBLE IDEAL BUT USEFUL BENCHMARK

One reasonable response to this indictment runs as follows: "The work cited is indeed suffused with values. It requires no great act of imagination to guess the value priorities, even political preferences, of the investigators. But it is a mistake to suppose that the work cited is any more value-laden than that of other investigators (Lebow, Jervis, Holsti, Axelrod, George, Sniderman, Carmines, . . . etc.). Value neutrality is impossible. We are all tainted and equally so."

This argument merits careful consideration. I agree that we are all tainted, but not equally so.

What grounds do we have for supposing that value-neutrality in political psychological research is impossible? At first glance, there appears to be an airtight logical argument which runs roughly as follows. Every science requires investigators to make a host of methodological and inferential decisions prior to hypothesis testing. Investigators have to decide how to frame the hypotheses to be tested, how to operationalize independent and dependent variables, and how to set thresholds of proof for accepting or rejecting hypotheses. In most of
science most of the time, these decisions stir little controversy. Investigators isolate the neurotransmitter of interest and test how it impairs or facilitates performance on various tasks. Aside from regular readers of *Nature* or *Science*, no one pays a lot of attention. Even the peaceful life of our hypothetical neuropsychologist can, however, be suddenly disrupted if it begins to appear that his or her work has implications for gender differences or interethnic or intraracial differences in cognitive performance. Once someone has been offended, all the routine methodological decisions that serve as background to hypothesis testing become charged with moral or political significance. Why did the investigator focus on neurotransmitter X rather than Y (perhaps his goal was to make invidious comparisons)? Why did he fail to take into account the interactive effects of different neurotransmitters? Why did he define cognitive performance in this way rather than that? Why did he use this statistical procedure rather than that? Once we start searching for potential sources of bias, we are not likely to be disappointed.

Much hypothesis testing in political psychology inevitably occurs in not only the academic arena, but in the political arena. Consider, for example, the well-known trade-off between Type I errors (rejecting the null hypothesis when it is true) and Type II errors (failing to reject the null hypothesis when it is false). Setting tolerance for Type I and Type II errors in research on symbolic racism or in research on the efficacy of deterrence in international politics becomes a consequential political act in itself. By operationalizing variables in certain ways, by selecting a certain sample, and by employing certain statistics and inferential procedures, the research inevitably takes a stand on the relative importance of minimizing Type I versus Type II errors. In the context of work on symbolic racism, this stand requires weighing the risk of concluding that symbolic racism drives public policy preferences, when it does not, versus the risk of concluding that symbolic racism is a weak or irrelevant causal force when it is actually a potent variable (Sniderman & Tetlock, 1986). In the context of work on deterrence in international relations, this stand requires weighing the risk of concluding that deterrent threats work when they do not, or the risk of concluding that threats do not work when they do (Tetlock, McGuire, & Mitchell, 1991).

In short, political psychology can be easily politicized.

It is one thing, however, to argue that values can easily influence inquiry and quite another to argue that values inevitably drive and determine the conclusions of inquiry. There are weak, moderate, and radical constructivists. The weak constructivists maintain that if we are careful and conscientious, and, most important, self-critical, we can minimize (even if never completely eradicate) the influence of values. The strong constructivists dismiss this position as positivist poppycock. They note that carefulness and conscientiousness and self-criticism in research are always in the service of some set of values. The key question is: Whose values are driving your research? The weak constructivists respond that if

the strong constructivists really believe what they say, they should quit pretending to do science and join the political party of their choice. I side with the weak constructivists here. Value neutrality is an impossible ideal, but it still remains a useful benchmark for assessing our research performance. Indeed, the price of abandoning value neutrality as an ideal is prohibitively steep: nothing less, I believe, than our collective credibility as a science.

Let's now consider some specific strategies that the scientific community can take both to monitor how pervasive political bias may be and to make efforts to minimize it. I identify four strategies here: (1) rigorous consistency in evaluating counterfactual claims; (2) embedding experiments in representative sample surveys to isolate causes of public opinion; (3) quantifying the qualitative; (4) vigilant search for counterexamples to the received wisdom.¹

(1) Rigorous Consistency in Evaluating Counterfactual Claims

History is not an experimental science; there are no control groups. We cannot go back in time and reenact what would have happened if Churchill had been prime minister at the time of Munich or if Gorbachev had not become General Secretary of the Soviet Communist Party in 1985. The control groups "exist" (if that is the right word) only in the imaginations of observers of the political scene (cf. Tetlock, 1991). We are forced to rely on highly speculative forms of counterfactual analysis: what would have happened if X had or had not occurred? It becomes tempting to allow one's preconceptions to fill in the missing "control-condition" data points. In this respect, counterfactual history is like a projective test. People often see what they want or expect to see. And what

¹Critical readers may suspect at this juncture that I have boxed myself into a corner. On the one hand, I have argued that value neutrality is impossible. On the other hand, I have criticized colleagues for allowing values to shape standards of evidence and proof, and I have recommended strategies for containing the influence of values on research. The critic is entitled to ask: given that values will inevitably play a role in inquiry, what counts as undue or excessive influence? Who decides?

It would be a mistake to try to deduce a precise answer from first principles. It is best to proceed with some commonsensical and rough-hewn rules of thumb. A useful starting point is to note that although most political psychologists are liberals (in the late-20th century meaning of that term), most are also uncomfortable with the subordination of science to political goals, even if they sympathize with those goals. It makes little sense to be claiming to be engaged in the scientific pursuit of causal understanding when the answers have been determined in advance or when certain ways of even posing questions are proscribed. Science presupposes a degree of open-mindedness that strong political commitments preclude. We should be prepared to acknowledge that deterrence may often work, that reassurance may often fail, that conservatives need not be racists, that groupthink occurs among our friends as well as our enemies and is, in any event, very imperfectly correlated with failure-success, and that styles of reasoning that resemble our own can appear strikingly maladaptive or even immoral in certain settings. A political psychology that merely echoes the received wisdom of the liberal wing of the Democratic Party (or any other orthodoxy) will not succeed by the standards of scientific endeavor. It will generate few controversies and even fewer surprising discoveries that enrich our understanding of human nature and politics.
people see tells us as much about the inner mental workings of the observer as it does about the external political workings of the world.

The debate between advocates of the conflict spiral and deterrence perspectives on the end of the Cold War illustrates the ease with which imaginative observers can posit counterfactuals that conveniently justify their original policy prescriptions. Many advocates of the conflict spiral view of the American-Soviet relationship warned that the Reagan administration was fueling the fires for a new and futile battle of the arms race (Deutsch, 1983; White, 1984). The legacy of the Reagan administration would be further deterioration in the U.S.-Soviet relationship: reinforcing self-fulfilling suspicions of each other’s intentions and strengthening the self-serving military-industrial complexes within both countries. The policy of “Peace through Strength,” and especially the strategically destabilizing Strategic Defense Initiative, were pushing us inexorably closer toward nuclear war, toward midnight on the doomsday clock on the cover of the Bulletin of the Atomic Scientists.

The predicted historical outcome did not occur. This nonoccurrence does not, of course, mean that conflict spiral theory has been falsified (although it does suggest that spiral theorists should specify more precisely than heretofore the conditions under which “upping the ante” in an arms race will lead not to escalation cycles, but rather to subtle and sober reassessment of goals by the other side—cf. Downs, 1991). The nonoccurrence did, however, stimulate an interesting counterfactual counterattack on partisans of the Reagan administration who sought to claim credit for both freeing the East Bloc and ending the Cold War. Many observers of the liberal spiral persuasion argued that the Gorbachev reforms were the result of internally initiated political and economic forces that had nothing to do with American defense policy. The same internal forces would have come into play whether President McGovern or Reagan had been in office. This counterfactual defense may or may not be correct, but it certainly provides a clear-cut example of theory-driven thinking. One’s theory initially leads one to expect a given level of causal variable X (aggressive deterrence) to produce outcome Y (heightened hostility). Instead, ~Y (reduced hostility) occurs. One responds by arguing that ~Y would have occurred regardless of level of causal variable X. The occurrence of Y has now, in retrospect, become irrelevant to the theory that inspired the original warning. 2

2In a recent article, White (1991) makes the surprising concession that the Reagan arms buildup may “in one way” have had a “very good effect”: By intensifying the pressure on the economically hard pressed Soviets to increase defense spending, the Reagan policies may have shocked the Soviet leadership into recognizing the futility of further military competition. Judging from all his writings, White probably did not intend to claim here that Reagan was either a necessary or sufficient cause of the end of the Cold War; at most, Reagan is granted a facilitative role. This concession reduces (although does not eliminate) the problem of inconsistency in standards for counterfactuals. However, it creates a new, perhaps bigger problem: does it follow that we should be grateful we did not adopt White’s recommended posture of “minimal deterrence” in the early 1980s because that would have given the Communist hardliners a much needed breathing spell in which to consolidate their hold on power?  

“Political Psychology, Politicized Psychology”

Conservative deterrence theorists are not in a much stronger epistemological position. Although many were eager to claim credit for the Cold War’s end, many also were on record defending a rigidly “essentialist” view of the Soviet Union which assigned virtually zero likelihood to the emergence of a Gorbachevian style of reformist leadership in a Soviet-style totalitarian system (recall the Jeanne Kirkpatrick distinction between reformable authoritarian and unreformable totalitarian regimes). Indeed, many went on record through the late 1980s to dismiss the significance of the movement toward freedom of expression and democracy in the East Bloc (in the alliterative phrase of one influential conservative, Gorbachev was a “neo-Stalinist in Gucci garb”).

In sum, policy advocates are often tempted to claim they know more than they do (or even possibly could) about the causes of historical events. This temptation was especially great in debates over American defense policy for two reasons: (a) the stakes were extraordinarily high and experts therefore found it especially dissonant to acknowledge ignorance (“Yes, I know the fate of humanity hinges on these decisions, but I still can’t offer you compelling grounds for making a decision. Why not toss a coin?”); (b) there were no good historical precedents for the nuclear stand-off between the superpowers. As Joseph Nye (1987) noted, much of what passed for knowledge about nuclear policy during the Cold War “rested upon elaborate counterfactual arguments, abstractions based on assumptions about rational actors, assumptions about the other nation’s unknown intentions, and simple intuitions.” Or, as Robert Levine (1991) noted in his history of arms control policy during the Cold War, policy debates were “ultimately theological in nature.” Virtually no one changed his or her mind because there was virtually no reason to do so. We had no evidence on which policies increased or decreased the likelihood of nuclear war. We learn from experience, and we have had no experience of nuclear war since the advent of the nuclear age in 1945. It was easy for the nuclear theologians to construct a dizzying variety of causal accounts for the nonoccurrence of a unique event.

I do not mean to imply that counterfactual arguments should play no role in political psychology. Counterfactuals are critical to our limited ability to learn from history (Breslauer & Tetlock, 1991). Counterfactual inferences should always, however, be subject to rigorous scrutiny. Specifically, I propose three tests for counterfactuals (tests that advocates of both spiral and deterrence positions typically failed throughout the Cold War):

(a) Have you applied the same standards of evidence to theory-consistent and theory-challenging counterfactuals? For example, have you been too quick to credit good outcomes to policies you advocate and too slow to give the same credit to policies you oppose? Do you generate counterfactual defenses when preferred policies are associated with bad outcomes (it would have happened anyway . . . ) but slight the same counterfactual possibilities when your adversary’s policies are associated with bad outcomes?
invoking violations of traditional values such as “the problem lies in the decline of the work ethic” or in “rampant sexual promiscuity.”

My principal quarrel with new racism researchers is their tendency to test the fusion thesis with “symbolic racism” scales that consist of items that deliberately mix ideologically principled and racially prejudiced grounds for opposing policies to assist blacks. In collaboration with Paul Sniderman and others, we have conducted a series of experiments embedded in representative sample surveys to provide fairer tests of the covert racism and fusion theses. The specific goals of these experiments have included: (a) to establish the conditions under which blacks are penalized because they are black; (b) to test directly for covert racism by (experimentally) supplying justifications that invoke violations of traditional values; (c) to determine the extent to which conservative or traditional values (variously operationalized) encourage racial double standards in judging who is entitled to government assistance. To achieve these goals, we used Computer-Assisted Telephone Interviewing (CATI) technology to manipulate experimentally the presumed race and sex of target beneficiaries of government aid as well as whether the target beneficiaries upheld or violated traditional values (good vs. poor work record, intact family vs. single parent).

The results suggest a more complex picture of the dynamics of racial policy reasoning than the symbolic racism position would lead one to expect. We find, for example, that conservatives are less likely to support government assistance for the needy, but they are not especially hostile to government programs designed to benefit black Americans. Indeed, even when we tried to incite symbolic racist responses experimentally by describing putative black beneficiaries of government programs as violating traditional values (undependable workers, single-parent families), conservatives failed to rise to the bait and refrained from being especially punitive toward black claimants. In fact, we find that conservatives are especially generous (by conservative standards) toward black claimants whom they believe are upholding traditional values by working hard and trying to make their own way (generosity that is paradoxically linked to the tendency of conservatives to be especially likely to harbor negative stereotypes of blacks and therefore to be surprised when presented with black stimulus persons who manifest strong support for traditional values). We also find support in this national survey for a long-standing hypothesis in the prejudice-and-politics literature: namely, evidence of double standards that put blacks at a competitive disadvantage in the struggle for government assistance and protection is most pronounced among the least educated (and therefore least familiar with ideological abstractions) in the general population sample.

It would, of course, be foolish to claim this study as the last word on symbolic racism or on racism in general. It is helpful, though, from a scientific point of view, to try to disentangle the causal roles of political ideology and racespecific attitudes in driving American thinking about government programs.
(3) Quantifying the Qualitative

Case studies play an indispensable role in political psychological research. There is no substitute for nuanced and contextually rich characterizations of how particular political actors felt, thought, and acted in particular situations at particular junctures in history. At their best, case studies check the overly ambitious theorist’s desire to assimilate new instances and to generalize into new domains. Ugly facts have been known to slay beautiful theories.

At their worst, case studies can be unfocused, loosely reasoned, vague, impressionistic, and self-serving. Although Alex George and others have made heroic efforts to formalize and systematize case study and comparative case studies methods (George & Smoke, 1974), most case study investigators are fiercely independent and resist anything that remotely smacks of regimentation. Case studies constitute a problematic database for our discipline for two mutually reinforcing sets of reasons:

(a) The additivity problem. Different authors often focus on different aspects of the same events. One author might trace decisions to the personality of the leader through a series of vivid anecdotes; a second author might stress the importance of factional in-fighting; a third might emphasize the failure of the group as a whole to monitor key economic trends. A reader might conclude that the authors disagree when they have completely complementary perspectives. They merely highlighted different aspects of a complex but shared reality. Moreover, the inferential problems are compounded when we try to compare case studies not of the same political leadership but of different leadership groups. We must now cope not only with theoretical and stylistic variations among researchers but also with highly idiosyncratic variations across time, cultures, and political systems. Here we confront the classic limitation of case studies. They may be wonderfully written, rigorously researched, and concisely conceptualized, but they don’t add up (Verba, 1967). We lack a systematic framework for accumulating insights across case studies;

(b) The susceptibility-to-bias problem. Case studies lack the usual precautions for minimizing intentional or unintentional error in the sampling and interpretation of evidence (precautions such as double-blind coding). In the typical situation, the solo investigator—who has a political or theoretical ax to grind—searches through the masses of historical evidence and decides which facts to highlight and which ones to downplay. The investigator also usually knows the outcome of each decision, so there is great potential for certainty of hindsight to color judgments of decision-making process. Finally, the investigator decides which historical cases to examine and which ones to exclude from his or her sample (again, much potential for selectivity and bias).

A group of collaborators and I recently developed a research method—the Group Dynamics Q-sort—that provides at least a partial solution to the additivity and susceptibility-to-bias problems that plague case studies in one area of political psychology: case studies of leadership group dynamics. The Q-sort consists of 100 statements that describe a wide range of structural and functional attributes of leadership groups. The instrument, for example, asks judges to rate the extent to which the group leader is insulated from criticism or exposed to a wide range of arguments; the extent to which the group is tolerant of dissent or punishes deviants severely; and the extent to which the group is operating under conditions of information overload and time pressure. We can address the additivity problem by circumscribing the conceptual universe of case studies of a decision-making episode, sampling case studies from this universe, and assigning research assistants the task of reading each case study and then rating the group portrayed in each study on the Group Dynamics Q-sort. Insofar as independent (theory-blind) readers agree on the nature of the group dynamics portrayed in each case study (high interrater reliability in Q sorts), we will have created a common metric for adding up case studies. We can now ask whether similar portraits of group dynamics emerge across case studies (high intertext reliability in Q-sort characterizations) and what the key points of agreement and disagreement are across case studies of the same leadership group. We can also gauge potential sources of errors and bias in case studies of group decision making. When only one case study leads us to the conclusion that the leadership group reached a decision in a rigid and defensive way and all other case studies suggest a reasonably flexible and open-minded decision process, we should seriously consider the possibilities of bias either in the single case or in the multiple case studies.

In a demonstration study, Tetlock, Peterson, McGuire, Chang, and Feld (1992) applied the Group Dynamics Q-sort to Janis’s (1982) case studies of groupthink as well as several independent case study interpretations of the same decision-making episodes. The cases examined ranged from the Chamberlain cabinet’s decision to appease Nazi Germany to the development of the Marshall Plan, the Truman decision to advance beyond the 38th parallel in the Korean War, the Kennedy decision to launch the Bay of Pigs invasion of Cuba, the Kennedy handling of the Cuban Missile Crisis, the Johnson decision to escalate American involvement in the Vietnam War, the Nixon Watergate decisions, the Ford Mayaguez decision, and the Carter Iran rescue decision. Our methodology was straightforward. We asked research assistants to read both the Janis case studies and the other authors’ case studies of each decision-making episode and then to rate the group portrayed in each case study on the 100 standardized dimensions of the Q-sort. We had multiple readers for each text, and we had multiple textual perspectives for each group.

Those of us who knew Irving Janis know that he did not pretend to be a neutral observer of the political scene. He believed that for much of the post-World War II period, the U.S. Government had far too low a threshold for using
force to achieve foreign policy objectives. But he also believed that the groupthink analysis captured important parts of the psychological reality behind a number of momentous 20th-century decisions. How can we tell whether he was right?

If Janis's classification of episodes into the categories of groupthink and vigilant decision making was indeed contaminated by his political values (a tendency to approve of the quality of group decision making when government leaders resisted the temptation to use force and a tendency to disapprove when government leaders resorted to force), then we should observe considerable divergence in Q-sort assessments derived from Janis's case studies and those derived from texts authored by a reasonably broad cross-section of historical observers.

What did we find? The short answer is that Janis's analysis holds up reasonably well. Substantial historical consensus indicates that the hypothesized cases of groupthink do indeed display many more symptoms of groupthink (in-group demands for loyalty, suppression of dissent, stereotyping of opponents) than do the hypothesized cases of vigilance (the Marshall Plan and Cuban Missile Crisis). The differences are resoundingly statistically significant. That said, however, we also find evidence of potential bias. Janis sees stronger and more consistent signs of groupthink in the historical cases he examined than do other historical observers. Q-sorts derived from reading Janis's case studies of groupthink episodes portray greater conformity pressure and greater rigidity in decision making than do Q-sorts derived from reading other authors' case studies of the same episodes. In addition, we find some historical cases that Janis classifies as evidence of groupthink that most other observers do not see as evidence of groupthink—in particular, the Ford administration Mayaguez rescue decision and the Carter administration Iran hostage decision. Although both decisions worked out badly, there is considerable evidence that decision makers were aware of the risks they were taking and made calculated trade-offs. These results do not decisively prove Janis to be wrong, but they do raise the possibility of either value-driven reasoning (Janis's moral disapproval of certain decisions colored his assessments of process) or certainty-of-hindsight reasoning (Janis was too quick to infer poor process from poor outcomes).

In brief, the groupthink Q-sort study shows how a widely used psychometric method can be deployed: (a) to quantify points of consensus and dissensus among historical observers of the same event; (b) to quantify patterns of similarity and difference across historical cases that previous investigators have placed in mutually exclusive theoretical categories (e.g., groupthink versus vigilance). Case studies can be added up. Potential examples of value-driven or theory-driven interpretation (outlier case studies) can be systematically identified.

(4) Search for Counterexamples to the Received Wisdom or Good Judgment

Let's turn to a fourth and final case where we need to monitor and check the influence of scientifically irrelevant value judgments: work on the cognitive structure underlying political ideology. In the early 1980s, I developed the value pluralism model of ideological reasoning to help explain a frequently recurring pattern that emerged in archival content analysis studies of political speeches and writings as well as in some experimental work. We observed that centrists and advocates of moderate left-wing causes tended, on average, to be more integratively complex in their policy reasoning than advocates of more extreme left-wing positions and both moderate and extreme conservative positions. This pattern of results holds up in many legislative bodies, including the United States Senate, the British and Canadian Houses of Commons, the Israeli Knesset, and the Italian Chamber of Deputies (see Putnam, 1971; Tetlock, 1984, 1986, 1989).

The tendency for moderates to be more integratively complex than extremists is not too surprising. After all, integrative complexity is defined in terms of evaluative differentiation and conceptual integration. Evaluative differentiation refers to the capacity and willingness to acknowledge the legitimacy or reasonableness of alternative perspectives on a problem. Conceptual integration refers to the capacity and willingness to develop integrative cognitions that specify ways of compromising or trading off conflicting perspectives and values. By virtue of their cognitive location on the political spectrum, moderates should be better positioned to see strengths and weaknesses in the arguments of the extremists. By virtue of the demands of political expediency and coalition building, moderates should be more motivated to strike deals with extremists at both ends of the continuum (and, by implication, to play extremists off against each other in political bidding for scarce resources).

What was puzzling was that the point of integrative complexity seemed, quite consistently, to be displaced to the left of center. Moderate social democrats tended to be more integratively complex than moderate conservatives. Although the old-fashioned ideologue hypothesis could not explain this aspect of the data, the value pluralism model was better positioned. This model posits that the point of maximum integrative complexity is displaced to the left because advocates of moderate left-wing causes are more likely than anyone else on the political spectrum to value fundamentally contradictory objectives: economic efficiency (let the invisible hand of free market competition work its magic) versus social equality (let's prevent excessive concentration of wealth and set a reasonably high safety net for the poor), deterrence (let's be strong enough to resist bullying by would-be foreign adversaries) versus reassurance (let's not strike so intimidating a pose that we precipitate conflict spirals), and economic growth (let the
economy expand to promote general prosperity) versus environmental protection (let's try to preserve fragile ecosystems and natural beauty). Drawing on classic work on cognitive consistency theory (Abelson, 1959), the value pluralism model maintains that people are under especially severe pressure to generate integratively complex cognitions when they place high importance on two values that prescribe contradictory courses of action. Here, people are most likely to acknowledge trade-offs, to try to specify ground rules for resolving trade-offs, and to try to understand why reasonable people view the same problem in starkly different ways. When value conflict is weaker (when one of the conflicting values is much higher in the value hierarchy than the other), people resort to simpler dissonance-reduction strategies such as denial and bolstering.

Where does this leave us? One possibility is that we have a reasonably value-neutral value pluralism model which simply attempts to explain a functional relationship between value conflict and cognitive coping responses. A number of observers have not, however, been content with value-neutral characterizations of these research findings. Some have gone on to argue that integratively complex reasoning is generally or even inherently superior — on cognitive or moral grounds — to integratively simple reasoning. (See Tetlock, Peterson, & Berry, 1993, for many examples.) One author of a personality textbook, for example, maintains that a “non-evaluative position” cannot be taken with respect to integrative complexity: “Here, complexity is superior to simplicity. The person who is integratively complex sees the world in more sophisticated terms, making more distinctions between various things and ideas (differentiation) and seeing more connections (integration)” (D. McAdams, 1990, pp. 552–553).

Following this author, should we conclude that moderate social democrats are, on average, “better thinkers” than left-wing extremists and moderate or extreme conservatives? Such a conclusion strikes me as premature, even a bit presumptuous and self-serving (given that most political psychologists tend to be liberal or social democratic in their sympathies). But how can we check this powerful human tendency to put a positive “value spin” on styles of thinking that correlate with views we happen to endorse? Here I can think of no research strategy better than looking long and hard for counterexamples in which the same style of thinking correlates with views that most of us find deeply objectionable or wrong-headed. So far, in our own research, we have been able to identify at least four such examples (two in archival work, two in experimental):

(a) The slavery debate in America of the 1850s (Tetlock, Armor, & Peterson, 1992). The major political viewpoints of the time can be roughly aligned along a left-right continuum: the radical Abolitionists (for whom slavery in any form was intolerable), the Free-Soil Republicans (who regarded slavery with deep distaste but were willing to tolerate it in the South as long as it did not expand into new territories), Buchanan Democrats (who adopted a much more conciliatory posture toward the slave-holding states and were willing to allow slavery to expand into new territories if local majorities so favored), and fire-eater advocates of slavery (who insisted that slaves be regarded as Constitutionally protected forms of property and that slave owners be allowed to take their property with them throughout the Union). Here we found that the point of maximum value conflict (between freedom and equality on the one hand, and states’ rights and property rights on the other) occurred somewhere between Free Soil Republicans and Buchanan Democrats. Maximum integrative complexity occurs among moderate Free-Soil Republicans (such as Abraham Lincoln). Few contemporary observers would be willing to argue that integratively complex politicians who were prepared to tolerate slavery into the indefinite future were either cognitively or morally superior to integratively simple politicians who saw slavery as an intolerable abridgement of fundamental Constitutional rights.

(b) The appeasement debate in Britain of the 1930s. (Tetlock & Tyler, in preparation). Neville Chamberlain had a substantially more integratively complex view of Nazi Germany and its intentions that did Winston Churchill. Chamberlain saw the Nazi regime as a difficult and disruptive force in Central Europe that could nonetheless be induced to moderate both its domestic and foreign policy to accommodate the norms of civilized states. Churchill had a starkly simplistic view of Nazi Germany (a “gangster state”) that led him to anticipate quite correctly the general trajectory of that nation’s domestic and foreign policy in the 1930s. By any reasonable standard of accuracy, Churchill was the more prescient judge of Nazi Germany.

(c) The debate over what constitutes acceptable risk in the Food and Drug Administration. Tetlock and Boettger (1994) carried out an experimental simulation in which we asked subjects to role play Food and Drug Administration regulators whose task was to decide on the admissibility of an anticoagulant drug into the U.S. pharmaceuticals market. The more integratively complex the subjects were in their thinking about the problem, the more likely they were to procrastinate (by requesting additional time for decision making) and to buckpass (by recommending that another government agency be consulted in the decision-making process). The complex subjects acted this way, moreover, even though the cost-benefit equation always indicated that the drug would save many more lives than it would jeopardize, and even though subjects were told that no new strong evidence was likely to arise in the permissible delayed action period, and that other government agencies did not possess information that the FDA lacked. Integrative complexity was correlated with decision-avoidance strategies that appear to have been motivated primarily by blame-avoidance.

(d) The dilution effect. Nisbett and Zukier have shown that people often lose confidence in the predictive power of genuinely diagnostic cues when those cues are embedded in sets of irrelevant or nondiagnostic cues. Building on this work, Tetlock and Boettger (1989) showed that integratively complex thinkers are unusually susceptible to this dilution effect. They try too hard to look for predic-
tive relevance in irrelevant information. Asked to predict the academic achievement of a student, for example, they focus not only on the diagnostic cue (the number of hours of studying per week) but also on the nondiagnostic cues (dating habits, frequency of tennis playing), making valiant but misguided efforts to integrate the latter cues into their predictions of overall academic potential (for example, specious inferences such as, “The guy is likely to be emotionally unstable because he has never dated anyone for longer than a couple of months. He is unlikely to be a good student” or “The guy sounds pretty athletic. He probably has a lot of energy and is a good student”). Complex thinkers use more information, but they are not necessarily more discriminating judges of the usefulness of information.

In short, we make a grave mistake when we allow our own political values and preferences to color our psychological characterizations of cognitive style. A cognitive style that in one context we applaud as balanced, judicious, and sophisticated we might deplore in another as weak, indecisive, and confused.

**CONCLUDING REMARKS**

This article is open to misinterpretation on several fronts. I conclude therefore by clarifying several aspects of the argument that I have advanced here.

It is obviously tempting to politicize political psychology. When I look back at the topics I have studied—the roles of deterrence and reassurance as causes of war and peace, learning in U.S. and Soviet foreign policy during the Cold War, the role of racism in shaping the racial policy preferences of the American public in the late 20th century, the role of cognitive and social processes in shaping the quality of decision making at both individual and group levels of analysis—I see a lot of value-laden terms. To preserve the peace, to be integratively complex, to learn, to engage in high quality decision making, and to be tolerant on issues of race are generally considered to be good things; to be racist or simple-minded is generally considered to be a bad thing. Skeptical observers are right to ask us what motivates us to pose such questions. Do we seek scientific knowledge of causal relationships? Or do we seek to advance certain moral or political causes by stigmatizing groups with whom we disagree (as racists who fail to learn, engage in groupthink, and cause wars) and applauding groups with whom we sympathize (as generous and fair-minded souls who do learn, engage in vigilant decision making, and preserve the peace)? These skeptics raise serious questions that merit serious responses. We should be candid about our motives as political psychologists. Very few of us, I suspect, are driven by purely epistemic motives or by purely partisan motives of policy advocacy. We are motivated, in part, by causal curiosity and in part by the desire to make the world a better place in which to live. And, being human, we don’t like to acknowledge that these goals occasionally conflict.

My own view is that epistemic and advocacy goals frequently collide. The most overt cases of politicization tend to occur when evidence of causality is particularly weak and the policy stakes are particularly high. For example, if I believe that irresponsible national leaders are pushing the world ever closer toward nuclear holocaust by overrelying on threats, it will strike me as irresponsible to insist that we scrupulously observe the standard norms of scientific evidence and proof. If I believe that racism in America has taken on new, more subtle, and ideologically sophisticated forms, it will strike me as irresponsible to rely on measures of crude, old-fashioned prejudice (irresponsible because such measures may provide a misleadingly upbeat view of the decline of racism in America). It is understandable that political psychologists as citizens often lend their voices to one or another political cause; it is less understandable when political psychologists (consciously or unconsciously) bend normal scientific standards of evidence and proof to advance those same causes (cf. Suedfeld & Tetlock, 1991).

Granted that the temptation to politicize our field is often strong, to what lengths should we go to resist temptation? I find it useful to frame this decision-making problem in value pluralism terms. If I attach high priority to the epistemic autonomy of science and low priority to policy activism, my choice is an easy one. I will argue that we should never bend standards of evidence and proof, no matter how morally inspiring the cause. It is not enough, moreover, just to admonish everyone to be fair-minded; we need to institutionalize procedures that apply the same methodological and logical standards to politically popular and unpopular hypotheses alike (no small order in an ideologically skewed subfield where most researchers and reviewers hold left-of-center values). If I hold the opposite value priorities, the choice will again be easy. I will argue for the “consciousness-raising” functions of political psychology and for an active role of the discipline in promoting desired social change. If I attach high importance to both epistemic and policy advocacy values, I will argue for some form of integratively complex compromise solution that will look neither attractive nor principled to the “extremists.”

I take an integratively simple stand on this question. It is generally a bad idea to dilute the goal of doing high-quality science with a host of additional moral and political goals. I believe that if we fail to institutionalize checks on the overt politicization of political psychological knowledge (most importantly, rigorous and even-handed peer review), we sacrifice our collective credibility as a science. We become just one more partisan voice clamoring for media atten-

---

*I invoke the parallel with research on the dilution effect in social judgment deliberately.
tion. Because we choose to investigate such controversial topics, political psychologists—more than most scientists—should continually ask themselves "Stockman's question" (named after the former director of the Office of Management and Budget in the first term of the Reagan administration): "Are we ideologues masquerading as scientists: Have we rigged the research dice in favor of our political agenda?" Moreover, we should ask that question not just because we fear what might happen if conservatives should, once again, gain control over the purse strings of our research budgets. We should ask these questions because we claim—in our journals, in our conversations with those who wield power—to represent a self-correcting scientific community.

Note: An earlier version of this paper was presented at the Society of Experimental Social Psychology in San Antonio, Texas, October 1992.

REFERENCES

Commentary
Burden of Proof

Paul M. Sniderman
Stanford University and Survey Research Center,
University of California, Berkeley

As against purportedly true-to-the-fact accounts of how science is not true-to-the-facts, I suggest that there is no place outside science from which one can judge science. From this, it follows that a concern for why a person has done an aspect of science is irrelevant to judging the validity of the science he or she has done. Political psychology runs into difficulty, I want to suggest, not because of the strength or mix of motives of those who do research in the field, but because of the weakness of the critical standards of those who evaluate it.

I am not aware of another psychologist who can match Philip Tetlock in range of coverage of politics or in fertility of conception. We have collaborated closely for more than a decade, and I mean to take this occasion to exhibit the spirit of our collaboration and provide a concrete example of how the practice of social science, so far from requiring consensus, can certainly tolerate and possibly benefit from differences of view.

Tetlock’s "Political Psychology or Politicized Psychology" tells a story which has no need for time subscripts—analyses of politics gone awry because of the politics of analysts. Without implying disagreement with his assertions of fact, I want to put a different gloss on the story; a different gloss because where we differ, so far as we differ, lies in our attitudes toward science as an enterprise and hence the nature of error within it. At the risk of over-drawing the differences between us, let me offer a thumbnail sketch of his position and mine. Tetlock sees the problem to be political moralism; I think it is cognitive simplism. He points a finger at those who do social science; I would call attention to those who evaluate it. He sees error as an inevitable and unfortunate blemish; I, in contrast, believe it is the best thing that scientists—even social scientists—have going for
them. And this string of differences between us cashes out in his pessimism and my optimism about social science.

"Weak constructivist" is the label Tetlock pins on his epistemic position. The job of the noun is to convey the assertion "that values can easily influence inquiry," with the task of the adjective being to retract, so far as possible, the assertion of the noun. Hence his characterization—"The weak constructivists maintain that if we are careful and conscientious, and, most important, self-critical we can minimize (even if never completely eradicate) the influence of values." (p. 9) Tetlock's causal-normative story divides into two—diagnostic and therapeutic. On the diagnosis side, political psychology goes awry because the political aims—values, concerns, biases, etc.—of investigators lead them to make any of many mistakes—ruling evidence in and out to suit the political argument they mean to advance, mismeasuring variables, misspecifying causal sequences, overgeneralizing and the like. And on the therapeutic side, the idea is to decrease the frequency of mistakes through investigators' becoming more self-aware and self-critical.

The adoption of better methods—experimental design and more rigorous measurement—is mainly what Tetlock has in mind, and this seems to me unexceptionable. It is, however, unexceptionable because it is good practice whatever an investigator's motives. In contrast, I think that Tetlock's presumption that it pays to attend to motives gets things wrong twice over. It offers a nontrivial risk of making things worse and poor odds for making them better. Let me take the two in turn.

Put into practice, Tetlock's argument represents an investment strategy. Better science will get done, he posits, if more effort and time is put into analyzing, discussing, and arguing why researchers do what they do. In contrast, I suspect the costs will outweigh the gains. Tetlock's argument represents, after all, a recommendation to put scarce resources (of several forms, but most especially time) into an activity—debating why an investigator is undertaking the investigation that he or she is conducting—when it is by no means obvious that we are especially competent as scientists in settling disagreements of this kind. And if it is set that the motivational inquiry that is being urged takes the form not of inter- but of intrapersonal analysis, then it seems to me the argument becomes merely hortatory since self-observance, being necessarily private rather than public, is unenforceable. Futility ought to be a decisive consideration, if not with respect to morality than at any rate with respect to science. There is not much to recommend requiring people to follow rules strictly if the rules are inherently unenforceable. And we can lose much in the process. If it becomes a legitimate part of social science to call into question why an investigator did what he or she did, the risk of politicizing political psychology—in a sense other and more serious than the one Tetlock has in mind—is surely nontrivial.

Ruminating on martyrdom, T. S. Eliot constricts the last temptation placed before Thomas Beckett—caught as Archbishop of Canterbury in conflict between his church and his king—as being Beckett's desire to do the right thing for the right reason. The temptation, to Eliot's way of thinking, is ruinously egoistic, distracting attention from the service of God to a self-interested (in two senses) absorption with merely human motives.

The temptation, I suspect, is as ruinous for science as for religion. To do the right thing has to suffice; it is irrelevant whether the act was done for the right reason. It misconstrues the problem to suggest, as Tetlock does, that political psychology gets itself in trouble because of the intrusion of improper motives in the gathering and weighing of evidence and can get itself out of trouble through motive monitoring, whether on an individual or collective basis. This whole focus on the motives of investigators, I want to suggest, gets wrong how science works, how it gets into difficulties, and above all, how it works its own way out of them.

The problem is not that political psychologists, because of the press of their concerns, values, and needs (including, though by no means limited to, their frankly political aims and biases) get things wrong, although they manifestly do. Rather, the problem is that when their research expresses or promises to advance a politically popular point of view it is not as likely to be held up to demanding critical standards. Where things go wrong for political science qua science is not in the play of motives in the design and conduct of research but in the modesty of intellectual standards in their evaluation.

Consider the size of errors. Mistakes are distributed along a continuum from egregious at one pole to subtle at the other. Politicized errors come in different sizes, but part of what I take Tetlock's argument to be is that the distribution tends to be skewed. The press of political aims increases the probability of egregious misattributions (particularly of validity and causal inference), he suggests. As a purely empirical matter, this strikes me as correct. Without suggesting that politicization cannot lead to subtle mistakes, his examples suggest that it favors egregious ones—mistakes that, when pointed out, are "howlers"—mistakes, that is, that are obvious once pointed out; mistakes, moreover, that obviously should have been pointed out on any close, critical examination.

Nor is it difficult, on reflection, to think why politicization should increase the likelihood of egregious errors. The effort to serve a larger public cause works against a willingness to put one's assertions to severe tests; and the less intense the critical self-scrutiny, the greater the likelihood of a howler.

This proposition ought to underwrite faith in social science. If politicization favored subtle errors, then the probabilities of detection and correction would plummet. But if politicization's effect is instead to increase the chances of the kinds of egregious errors that Tetlock cites—for example, the argument that because symbolic racism is defined to be a conjunction of prejudice and values, it is then proper to measure it with "questions that deliberately mix racist senti-
ments and traditional values” (Kinder, 1986, p. 156, italics added) - then it tends, precisely because it is self-defeating, to be sustaining of the larger enterprise. Who can seriously debate that a measure should be ambiguous?

Let me say only a word about why a high value ought to be attached to egregious errors particularly at this stage in the development of social science. An egregious error, whether easier to commit than a subtle one or not, is very much easier to detect. And though it may be too much to ask that a person deeply committed to a research project should avoid making a mistake, egregious or otherwise, it is not too much to ask that a critic, scrutinizing the research, should spot the mistake especially if it is egregious. And since the immediately crucial task in promoting social science is to raise the standards of the people who evaluate it, it is fortunate that mistakes can frequently be easy to expose. A generous schedule of reinforcement should build a stronger habit of critical scrutiny.

Both more and less divide Tetlock from myself than may appear evident at first sight, and one way to capture both is this. Tetlock raises the question of the judgments to be pronounced on political psychology, given the claims of (weak) constructivism. In contrast, the question I believe should be raised is: What judgment should be pronounced on constructivism, given the claims of naturalism?

Constructivism, though available in different degrees and formulations, is centered on three claims: that science is not rule-governed; that scientific disputes cannot be settled by empirical evidence; and that “nature” does not determine science; instead . . . “the social behavior of the scientists in the laboratory determines how the laws of nature are defined” (Cole, 1992, p. 5). As against this, my preference is for naturalism. I take naturalism to consist, at its core, in the claim that there is no position outside—and superior to—science, from which the procedures and claims of science can be judged. As Quine (1981, p. 21) once put it, naturalism consists in “the recognition that it is within science itself, and not in some prior philosophy, that reality is to be identified and described.”

By way of confession, I want to acknowledge reading the work of constructivists like Latour and Cole, and finding myself both engaged and enlightened. But they can manage this for exactly the same reasons that historians—or literary critics—can be insightful and instructive: because intelligent people, writing about the behavior of (sometimes) intelligent people, can say much that is insightful and instructive. The sociologist of science is practicing a form of historical analysis, and as a systematic statement of reality can claim as much as history itself can claim—which is to say, not much. Partly, I have in mind here the self-contradictory character of the claim of most forms of constructivism—namely, that it is offering a true-to-the-facts account of how science is not itself true-to-the-facts. But, still more, I have in mind the presumptuousness of supposing that the “conclusions” presented by constructivists provide a superior vantage point from which the claims of scientists to truth can be weighed. There is, surely, something amusing in the latest run of fashion having installed philosophy and history as branches of knowledge rightly placed to pronounce judgments on the validity of science. There is no vantage point outside science—whether in history, sociology or philosophy—from which one can make a judgment about whether science (in its quite various manifestations) is being well-done and whether it should be differently done. One can only put things right from within science, by doing it.

The actual practice of science reflects the fact that its practitioners are actual people. But as William James might put it, the question is the cash value of this observation. Constructivism takes the lesson to be that science will inevitably get things wrong. In contrast, from the point of view of naturalism, the issue is not whether people fully embody abstract ideals but rather whether their behavior will more closely approach ideals if they aspire to them. On this second point, Bauer (1992), in my judgment, has hit the nail on the head, observing that:

Those who hold ideals, no matter that they are unattainable, are likely to behave more in accord with them than will people who do not hold those ideals. . . . That human beings cannot by nature be entirely objective does not render objectivity an unworthy ideal. (p. 39)

There may not be a scientific method: but there is a scientific attitude, and it uniquely combines modesty and confidence, rejecting the quest for (and claim to) certainty, acknowledging the imperfection of the enterprise at any moment (Pierce’s “contrite fallibilism”), insisting all the same on science’s power to be self-correcting. Hence Quine’s (1960) famous rendition of Neurath’s classic metaphor: “Neurath has likened science to a boat which, if we are to rebuild it, we must rebuild plank by plank while staying afloat in it” (p. 3).

REFERENCES

Ideological Bias in Political Psychology: The View from Scientific Hell

David O. Sears
University of California, Los Angeles

Tetlock asserts that the work of Sears and Kinder on symbolic racism is “deeply politicized,” which has “profoundly shaped standards of evidence and proof.” He is particularly concerned that this research automatically indicts conservatives as racists. Contrary to his presupposition, in 10 empirical studies, conducted over more than two decades of research, and using a variety of different research strategies, we have assessed the independent roles of these two predispositions. We have most often found independent effects of both racism and conservatism on whites’ opposition to racial policies, although racism generally has had stronger effects. In this response, I indicate that only a highly selective review of our work would merit such a serious charge. As a general matter, I propose that self-consciously adversarial procedures, along with, rather than solely relying on, the oft-abused ideal of “value neutrality,” may best promote credible scientific research on politically controversial topics. I would also urge adherence to consensual standards of evidence, not just in word, but in deed.

KEY WORDS: racial policies; symbolic racism; racism; conservatism; political psychology; political controversy; value neutrality

Hell: A place or state of misery, torment, or wickedness; a place or state of turmoil, disorder, or destruction; a cause of torment, tumult, or havoc.
Heaven: The dwelling place of the Deity; a place or condition or period of utmost happiness, comfort, or delight; perfect felicity or contentment; a sublime or exalted condition; a transcendental cosmos or domain.
—Webster’s Third New International Dictionary

Philip Tetlock, in his “Political Psychology or Politicized Psychology: Is the Road to Scientific Hell Paved with Good Moral Intentions,” says that we are doomed to “scientific hell” when “researchers . . . feel so passionately about a cause . . . that those passions influence key methodological and conceptual de-
cisions in research programs. As a result, political psychology becomes politi-
cized,” leading to “the complete collapse of our credibility as a science. . . .
Thoughtful outsiders cease to look upon us as scientists and see us rather as
political partisans of one stripe or another.” “A political psychology that merely
echoes the received wisdom of the liberal wing of the Democratic Party (or any
other orthodoxy) will not succeed by the standards of scientific endeavor. It will
generate few controversies and even fewer surprising discoveries that enrich our
understanding of human nature and politics.” In other words, no one will believe
us, but even worse, we will be boring.

He then turns to “two specific examples of how political psychology can be
deeply politicized: Ralph White’s work on the causes of war and peace and David
Sears and Donald Kinder’s research program on symbolic racism.” There is in
Tetlock’s eyes, apparently, a strong political agenda here, in which the “morals-
political values of the investigators appear to have profoundly shaped standards
of evidence and proof in testing competing hypotheses,” and “political psycholo-
gists (consciously or unconsciously) bend normal standards of evidence and
proof to advance those same [political] causes.”

These are harsh comments. I will speak first to his analysis of our work on
public opinion about racial issues, which strikes me as surprisingly partial, both
poorly informed and, putting it delicately, “theory-driven.” Then I will address
his assumptions about what good science is and how one’s own political values
fit into that; here I distinguish between promoting a theory and promoting one’s
own political values. I propose a focus on the former, with a more explicitly
adversarial version of the “value-neutral” strategy than he espouses.

POLITICIZED RESEARCH?

His specific complaint is that “because Sears and Kinder have defined
symbolic racism as a fusion of antiblack affect and either traditional or conserva-
tive values, they have often relied on survey questions that deliberately confound
or mix these two elements” (p. 512). This research strategy “labels people who
object to busing or affirmative action on purely ideological (race neutral) grounds
as racist by definition.” Instead, one should “try to disentangle the causal roles
of political ideology and race-specific attitudes in driving American thinking about
government programs.” This argument repeats portions of his earlier critiques
co-authored with Paul Sniderman (1986a, 1986b).

He has one part of our conclusion right. Much of the research my colleagues
and I have published for over 20 years has concluded that racism has signifi-
cantly influenced white Americans’ opposition to policies intended to ameliorate
racial inequality (such as busing or affirmative action) or to black candidates
(such as Tom Bradley or Jesse Jackson). But the role of nonracial conservatism

was, from the beginning, also a central question for us. Does racism capture all
its effects, as Tom Emsell has argued (“All politics is racial”)? No, on the
contrary, we have consistently argued that both racism and ideological conserva-
tivism contribute, independently, to whites’ policy and candidate preferences.

A third question we have taken up is the form that contemporary racism
takes today. We argue that there is a “new” form of racism, which we called
“symbolic racism,” merging antiblack affect and traditional values. But the exact
form of contemporary racism is a separate issue from the question of whether
policy and candidate preferences have a racist component (however racism mani-
ests itself) or whether nonracial attitudes like general ideology make a contribu-
tion independent of racism. Much of our published work, and much of the work
he discusses, addresses the relative roles of racism and ideology, ignoring any
possible “new” form of racism. Similarly, my response here addresses the more
general issue of a racist component in policy and candidate evaluations, and
ignores the separate question of “symbolic racism.” For this latter, see especially

Tetlock correctly indicates that in some studies we have predicted such
opposition from items combining racial antagonism with nonracial traditional
values (e.g., Kinder & Sears, 1981; Sears & Allen, 1984). He credits the rela-
tionships we report as being “intuitively convincing.” But he complains that such
correlations could result from policy opposition among nonracist, principled
conservatives, without any necessary role of racism. This is an obviously plausi-
ble alternative explanation; indeed, so obviously plausible that we have ad-
dressed it from the very beginning in the four following ways, all of which he
ignores.

(1) One strategy was to control statistically on general political ideology
after correlating such “mixed” items with our dependent variables (Sears &
Kinder, 1971, concerning voting against Tom Bradley for Los Angeles mayor;
Sears, Hensler, and Sears, 1979, antibusing attitudes; Sears, Lau, Tyler, &
Allen, 1980, busing and “law and order”; Sears & Citrin, 1982, support for the
California tax revolt; and Sears, Citrin, & Kosterman, 1987, evaluations of Jesse
Jackson). We found in these cases that both racial attitudes and nonracial ideolo-
gy made independent contributions, though racial attitudes were in each case
considerably more influential.

Since this strategy was at the heart of our approach from our first study, and
Tetlock ignores this extensive set of analyses, it is perhaps worth quoting from
our original publication on this point, published over 20 years ago:

Since racism and conservatism are so closely related and since they apparently derive
from the same social background, a natural suspicion is that they are, in reality, one and
the same. Liberals have often alleged that there is no difference, e.g., that the “freedom of
choice” rationale for opposing fair housing is merely a cover up for racism. The question
is whether conservatism is anything more than that.
The answer is that conservatism most certainly is something in addition to a nervous response to racism. Even when symbolic racism is controlled, liberalism-conservatism was highly correlated with candidate preference in the mayoral election. Racism and ideological conservatism represent a real one-two punch in producing conservative voting today in southern California. They have cumulative effects.

Racism seems to be the more important political issue, however. On nonracial issues, conservatives of both parties resemble each other closely, as do liberals of both parties. On racial issues, however, the liberal Democrats stand alone (Sears & Kinder, 1971, pp. 76, 78, 83, 84).

This technique of statistical control is well-established. One would have to ignore these findings to maintain that our measures of racial attitudes are really just measures of conservatism and that that alone explains opposition to housing, black candidates, and so on.

(2) A second strategy has been to test both the effects of racial affect by itself (as measured in the “feeling thermometer”, not intermixed with nonracial themes), and the effects of such nonracial attitudes as party identification or political ideology (Sears et al., 1987; and Sears, 1988, on evaluations of Jackson; Sears, 1988, on racial policy; and Sears & Kosterman, 1991, on a wide variety of racial policies and candidates). Even this direct and indeed rather bald assessment of racial attitudes yields significant effects of them. Such analyses also generally yielded significant independent effects of political ideology. But more than conservatism is involved, and part of it, at least, is reflected in racial attitudes.

(3) A third strategy has been to test the effects of racial affect and symbolic racism and nonracial predispositions (party identification and ideology) on racial policy preferences all at once (Sears & Kosterman, 1991). In this case the independent effects of nonracial predispositions are much attenuated, though still statistically significant on occasion.

(4) A fourth strategy has been to compare the effects of racial attitudes on whites’ opposition to liberal black candidates (such as Jackson or Bradley), as opposed to their effects on opposition to liberal white candidates (such as Walter Mondale or Jerry Brown, among others: Sears et al., 1987; Sears, 1988; Citrin, Green, & Sears, 1990). Here the results are a bit more mixed. Contests involving black candidates all attracted significant racial-attitude effects, whereas those involving white liberals did in some but not all cases (for example, Brown-Wilson did, but Mondale-Reagan did not). Ideology always had strong independent effects in races involving white liberals, and usually but not uniformly in those involving black candidates (for instance, not with Jackson in 1984; see Sears, 1988).

In these latter studies, then, the external stimulus is important in whether or not racial attitudes (however labeled) are engaged. Usually, racial stimuli evoke racial attitudes. Racism is sometimes engaged even where the cue is not overtly racial, as in the tax revolt. But not all liberal-conservative contests evoke racial attitudes.

Throughout this research, then, racism consistently does predict whites’ opposition to racial policies and black candidates, even with various methods of measuring this concept and with some care to controlling on such nonracial attitudes as ideology. Tetlock pleads for efforts “to try to disentangle the distinctive sources of resistance to policies to assist various minorities and to set up one’s empirical procedures so that conservatives are not automatically guilty of racism,” and then puts forward his current research with Sniderman as the first effort to do so. But this is precisely what we have been doing for 20 years, using four rather different research strategies, and that work has appeared in most of the publications he cites.

Using the term “scientific hell” may have been intended to be cute, or ironic, but it is pretty strong. The language throughout strikes me as excessive, though more contemptuous of Ralph White than of my colleagues and myself. And the review overlooks extensive data analyses on what seems to be the very focus of its main concern, focusing only on the one aspect of it that fits its conclusion, that we indict conservatives as “automatically guilty of racism.” I will leave it for others to speculate about the reasons for the selectivity of such a review.

DOING CREDIBLE SCIENCE ON CONTROVERSIAL ISSUES

Let me now turn to the more general issue of this Symposium: how can political psychologists do research on matters that are fraught with political controversy without getting written off as mere partisan pleaders? To start with, we need to contrast two different issues. The first is whether or not one’s own political predispositions bias one’s research. The second is whether we begin with strong theories that organize our empirical work or start with data and induce generalizations from them. In political psychology, these two issues tend not to be as distinct as they are in some other fields (though we are scarcely alone in the social sciences in facing this problem).

Tetlock addresses only the former. He begins with two premises that I would wholeheartedly agree with, namely that “completely value-neutral political psychology is impossible” and that “our collective credibility as a science depends on self-critical efforts to monitor and minimize the influence of scientifically irrelevant values on inquiry.”

He then contrasts two versions of a disbelief in “value-neutral” social science. He sides with the “weak constructivists,” who try to get as close to “value neutral” theories and research as possible and to construct a variety of safeguards
to detect and check the effects of political bias on research. I think these safeguards are thoughtfully chosen, in general, and well worth heeding. He sees the major alternative to his approach as "strong constructivism" which rejects positivism and thinks that all such efforts are "in the service of some set of values." The "weak constructivists" should, in his view, tell the "strong constructivists" that "if they really believe what they say, they should stop pretending to do science, and join the political party of their choice."

Given just these two choices, I too would side with the weak constructivists. Indeed, some years ago, in publishing our empirical research on the earlier Los Angeles riots, my colleague John McConahay and I expressed very similar sentiments:

Most of what has been written about the [ghetto] riots has been political propaganda thinly disguised as scholarship, with little effort to test it against systematic empirical data.

We too have strong moral convictions about race relations. We have tried to protect ourselves against bias in three ways: we have tried to test our theories as rigorously as we could with the best empirical data we could gather; we have tried to formulate alternative theories that were as plausible as possible, and give them as fair a test as we could; and we have tried to be honest about the discrepancies between our theories and the data. We regard ourselves as scholars rather than as pamphleteers or activists. There is no way that we can completely screen out the influence of our biases on our work. Hopefully, we have presented enough of the data so that others can judge our work fairly." (Sears & McConahay, 1973, p. ix)

But let me develop this line of thinking a little further, and in so doing, offer a third alternative. Given the intense conflicts of values and interests in much of politics, the legal system, rather than traditional natural science, may be the more appropriate place to look for models of dispute resolution. A useful distinction is Thibaut and Walker's (1975), between the "inquisitorial" (or "fact-finding judge") and "adversarial" models.

The Tetlock "weak constructivist" approach parallels the former, in which all evidence is placed before an impartial judge, who then is called upon to make a dispassionate decision. Everyone knows that no judge is completely dispassionate or disinterested but the intention is to act as disinterestedly as possible.

An alternative comes closer to the adversary system on which the American legal system is modeled. Here the protagonists each take strong positions promoting their own view, presenting the evidence from that perspective. Both protagonists and spectators are bound by consensual rules of evidence. Over time, it is felt, the truth will out, and more clearly so for having had the best possible advocates passionately arguing each position.

This approach better fits the findings of contemporary social psychologists about how humans deal with controversy. Social psychologists have assembled a wide variety of well-understood phenomena under the general observation that human beings often engage in "theory-based processing," the tendency to interpret incoming data in terms of our prior cognitive structures (whether described as schemas, scripts, stereotypes, prejudices, attitudes, or the prior structure Tetlock singles out—partisan political preferences). In life as in science, the data never completely "speak for themselves." The "blooming, buzzing confusion" of life would overwhelm us if we let it.

One of the longstanding debates within psychology has revolved around the extent to which these organizing structures facilitate scientific progress, helping us to simplify and interpret a complex social world, or impede it by blinding us to the data and distorting our reading of it. The weak constructivist position resembles the inductive answer to this dilemma: let the data speak for themselves, and dispassionate scientists will recognize the message. The adversarial approach more closely resembles the hypothetico-deductive model, in which theory drives both data gathering and data analysis.

Social and personality psychology have had many examples of both models. In personality psychology, the Minnesota "dust-bowl empiricists" simply looked for correlates of various disorders in the MMPI, without theoretical guidance. In contrast, Hall and Lindzey (1957) in their book Theories of Personality, advocated that:

The notion of a general synthesis or integration usually communicates to the student the need to be cautious, take all points of view into consideration, and avoid emotional involvement with a particular, one-sided position. Before embracing any particular theory let us compare it with others, see that it is just as good in all respects, and read what the critics have had to say about it. Contrary to this conception, we recommend strongly that the student should, once he has surveyed the available theories of personality, adopt an intolerant and affectionate acceptance of a particular theoretical position without reservation... It will dictate problems to the devotee and stimulate him to do research but it will not do the same for the cool and detached observer... let him immerse himself in one theory of personality. Wallow in it, revel in it, absorb it, learn it thoroughly, and think that it is the best possible way to conceive of behavior.

In social psychology, Leon Festinger personified this adversarial approach, by making "non-obvious" predictions from strong theoretical predilections and arguing the strength of confirming data in a passionate, partisan manner. Nisbett and Ross's (1980) passionate advocacy of a cognitive approach to social perception is another example; they push it as far as it will go (and for my money, a little further, but we are all better off for the effort).

But if we adopt this more adversarial, theory-driven approach, how can the data be honestly judged? Will investigators not constantly bias their data collection and analysis to support their own theories? Well, juries in the American system listen to both sides, each presumably ably represented, and decide which is more plausible, according to standard rules of evidence. In the case of social science, we have not only a comparably strong and consensual body of rules of evidence to guide us, but even better, we have the critical test of replicability: which view bears up best under both literal and conceptual replication?
MEASURING REAL PASSIONS, OR NON-ATTITUDES?

It is easy to adopt the cool and lofty position of the dispassionate scientist, letting the scientific cards fall where they may. But it may not produce good research, as Hall and Lindzey suggest. So Tetlock does not like our use of "survey questions that deliberately confound or mix these two elements" (anti-black affect and either traditional or conservative values). His research, in contrast, presents subjects with short and abstract vignettes describing a hypothetical unemployed person, varying the person's race, gender, marital and parental status, and dependability (Sniderman et al., 1991). This is laudable in terms of experimental control, but like any example of standard experimental social psychology it is susceptible to the usual threats to internal validity, such as demand characteristics, social desirability biases, and manipulations that don't "take."

More important, from my perspective, do antiseptic, dispassionate, hypothetical vignettes really capture the tensions and conflicts and emotions that Americans have about racial issues today? There is a long literature in social psychology about the distortions introduced into the results by use of such hypothetical experimental tasks (Freedman, 1969; Hovland, 1959; Sears, 1966). It is clear which side I am on: first capture the human reality one cares about. Then worry about experimental and/or statistical controls. Racism is not pretty. Perhaps we need to get in and muck around with the passions and hatreds of real people concerning real issues. I would advocate using items whose wording maps onto the actual thoughts and feelings of ordinary people; otherwise we risk measuring "nonattitudes." If that means using items that are a little messy, and trying statistically to straighten things out a posteriori, so be it. People are a little messy too.

In conclusion, how does scientific heaven look from the vantage point of "scientific hell"? Well, heaven, as always, has white fluffy clouds, soft music is playing, its denizens are dressed in white, flowing robes, they are characterized by gentleness and selflessness and purity, they seem to have no human passions other than a selfless caring for others. Hell of course is characterized by dark, red, noisy stench, wherein its inhabitants are continually trying to gain some personal advantage with no regard for others or for truth or any kind of moral values; selfish and short-sighted, they are.

It is appealing to try to get the stench of hellfire out of our science, and hear only the gentle heavenly choir. But we are investigators (and readers) who have strongly "theory-driven" views of the world, whether the "theories" are framed in conventional scientific terms or in partisan political terms. Is it best to hide our own biases and pretend they do not exist, leaving it to others to wonder vaguely why we always seem to come out on one side or the other? Or should we try to be up front about them, as best we can; follow consensual rules of evidence as best we can; and leave it to others to judge whether our work is replicable, best accounts for the phenomenon, and so on?

To return to the theological metaphor, clinicians do not advocate that we adopt self-images as pure and wholesome and without sin; they advocate self-knowledge and acceptance of the evil within us. And being upfront about it with others (as in our book on ghetto riots, which began with a statement about our political proclivities at the time, preface the plea to let the issues be settled on the basis of the empirical data at hand, as quoted above [see Sears & McConahay, 1973, p. ix]). It is conventional for us to be explicit about our theoretical preferences but unhappily less so about our political preferences. Then let the jury decide, informed about where we are coming from. That seems to me far healthier than cloaking our own feelings in a pretense of scientific objectivity, while ignoring a ream of scientific evidence we happen to find distasteful.

ACKNOWLEDGMENTS

This paper was prepared while the author was a Fellow at the Center for Advanced Study in the Behavioral Sciences. I am grateful for the comments of Jack Citrin and Šule Ozler on an earlier draft, and for the financial support provided by National Science Foundation grant SES-9022192.

REFERENCES

Oh, Ye of Little Faith: Philip Tetlock’s Road to Hell

Brian Kroeger and Virginia Sapiro
University of Wisconsin-Madison

The truth of this is clear. For if it is asked why some witches will not confess the truth under even the greatest tortures, while others readily confess their crimes when they are questioned, . . . the reason is as follows. It may truly be said that, when it is not due to a Divine impulse conveyed through a holy Angel that a witch is made to confess the truth and abandon the spell of silence, then it is due to the devil whether she preserves silence or confesses her crimes. The former is the case with those whom he knows to have denied the Faith both with their lips and in their hearts, and also to have given him their homage; for he is sure of their constancy. But in the latter case he withdraws his protection, since he knows that they are of no profit to him (Malleus Maleficarum, 230).

In order, then, that the Judges both ecclesiastical and civil may have a ready knowledge of the methods of trying, judging, and sentencing in these cases, we shall proceed under three main heads. First, the method of initiating a process concerning matters of the faith; second, the method of proceeding with the trial; and third, the method of bringing it to a conclusion and passing sentence on witches. (Malleus Maleficarum, 430).

We all enjoy a deftly employed metaphor, and Philip Tetlock’s lesson on moral intentions, science and politics provokes us to beget one more iteration (Tetlock, this issue). That we disagree in important respects with the judgment, diagnosis, some of the articles of faith and, indeed, the method of revelation will become clear shortly. But basic to our disagreements is our objection to the metaphor itself. Within the first paragraphs we find moral intentions, temptations, forbidden fruit, passions, and the virtual Armageddon of scientific hell. As we read the essay, its metaphorical frame is not merely a clever entree into a difficult and important discussion; it is an indication of an epistemological leap of faith we cannot take.

We begin with our basic agreements, the better to clarify the points at which we diverge. Professor Tetlock’s overall project is a most worthwhile venture. Political psychology as a community has not yet adequately engaged in serious and necessary debate about the epistemology and practices (methodological and otherwise) of our discipline. Our interdisciplinary and international character

1The Malleus Maleficarum is the 1486 handbook on witchcraft, its character, and its elimination, as used by the Inquisition.
I. THE ANGELS AND THE WITCHES

Let us begin with a brief critique of Tetlock’s method of argument. It is too personal, detracting from the main argument, and it violates the very canons of science Tetlock advocates. The general point—that scholars’ political or moral viewpoints often shape their observations and conclusions—is rarely a subject for contentious debate any more. But extended criticisms of a few scholars, especially when the criticisms entail attributions of motive and causation for which there is insufficient direct evidence or consideration of alternative hypotheses, makes much of the body of this article regrettable. Aside about other authors (as in the footnote about Ralph White’s “surprising concession” [confession?]) also seem misguided. We will not participate in the debate over whether these individuals have committed cardinal sins against science, although we do detect a series of venial sins against the rules of scholarly civility. Henceforth we concentrate on the worthwhile broader issues raised by the article.

II. DEFINING SCHOLARLY HEAVEN AND HELL

According to Tetlock, scientific hell is the “complete collapse of our credibility as a science,” when “our powers of persuasion are limited to those who were already predisposed to agree with us,” and when “thoughtful outsiders cease to look upon us as scientists and see us rather as political partisans of one stripe or another.” Thus, the road to hell is paved with two elements requiring clearer distinction: (1) the impact of one’s values, expectations, and desires on the conclusions one reaches in scholarship, and (2) the impact of a scholar’s values on the conclusions as seen by various “outsiders.” These are different problems, and any solutions to them will be quite distinct. The former question refers to the criteria we apply to our own intellectual behavior (we shall have much more to say about this topic later). The latter relates to how our research is legitimated within our scholarly community (here, other political psychologists), but more importantly, to how the community of political psychology appears to “outsiders.” For Tetlock, scientific hell is caused by individual scientists’ insufficient scientific behavior, which has the effect of placing those researchers in a version of hell within the political psychology community. But a more serious potential effect of these individuals’ actions is to damn the community as a whole to descent into the shadowy nether realms of “nonscience.” For Tetlock, outsiders have the power to issue this ultimate punishment for transgressions of the canons of science.

Regarding this second aspect of scientific hell, we begin by agreeing with the implied point: outsiders indeed bestow credibility upon us. But who exactly these outsiders are is an important question for sociologists of knowledge. Tetlock never identifies them specifically, but among those relevant to us in different ways are other political psychologists who do not share our specific methods or interests; other social scientists who are not political psychologists; other academics outside the social sciences, such as natural scientists and humanities scholars; and various publics, bureaucrats, politicians, and professionals who employ us, fund us, supply us with fresh graduate students, set policy priorities, and read and apply our work.

The basis of our credibility varies among these different groups, and not just because some of them “know” science and some do not. The credibility of political psychologists is a more complicated sociological and social-psychological problem than Tetlock admits. He assumes that “outsiders” value roughly the same canons and practices of science as he does, to the same degree, and in the same order of priority. But consider that much of the broader scientific community, especially in the natural sciences, has never thought much of the social sciences as sciences even in their more “positivistic” incarnations.

Returning now to the first paving stone of the road to hell, the relationship among individual values, motives, and scholarly research, we agree with Tetlock.

*Perhaps this is scientific purgatory.

*The reader should not make the mistake of dismissing Tetlock’s and our use of metaphor as idle frivolity. The point of using the metaphor of “scientific hell” is precisely that hell is the ultimate imaginable punishment. Tetlock is not discussing “scientific weakness” or “scientific immaturity” te.g., the idea that the social sciences “just haven’t found their Copernicus yet”) but rather the potential that political psychology as a whole will be banished from the lofty realms of “science” entirely because of the irresponsible scientific sins of a part of the community.
that it is impossible to do entirely value-free research, or to eliminate the influence of “scientifically irrelevant” values. Tetlock’s road to scientific hell begins a bit further down the garden path of values, namely when individual scientists allow their “political passions to trump normal scientific standards of evidence and proof.”

III. VALUES AS ACADEMIC SINS AND VIRTUES

The individual scientist’s actions are tied back into this argument when we turn to the problems inherent in defining “science” and “scientifically irrelevant values.” Tetlock’s comments raise three related sets of issues about the meaning of “science” and “scientific criteria.” He alludes to the first important issue in a footnote, asking “what counts as undue or excessive influence [of values]? Who decides?” He suggests an unsatisfyingly commonsense imperative: Leave yourself open to discovery of the unexpected or undesired. As a political psychologist of a cognitive bent, Tetlock knows that, while true, this statement offers little guidance. It does not answer the question of who decides, except via the implicit command of “you must police yourself.” This fails to make the crucial connection between the individual’s actions and the fate of the community. Moreover, because Tetlock’s emphasis is on the nature of good science, not just good scholarship, we must point out that being open to the unexpected or undesired is not a uniquely scientific intellectual value. Our various training in history, philosophy, and even the literary and graphic arts included similar exhortations.

A second question is, What are the “scientifically irrelevant values” that should not intrude (too much) on scientific inquiry? Who decides what they are? For that matter, what are the scientifically relevant values that should intrude on science? We invert the question because it is important to remember that science is defined by certain shaping values. The “good” values enumerated or implied in Tetlock’s paper include the valuation of cognition and observation over passion, desire, or intuition in research; the value of maintaining credibility for the enterprise of social science as a science; the valuation of greater rather than lesser control in research (as in the fondness for experimentation); and the value of clarity and precision. Values are crucial to science.

The more difficult problem is indeed identifying the values that are irrelevant to science. Let us cut straight to the difficult cases. We find plausible Alfred North Whitehead’s argument that scientists can find their drive for science (and science performed within the conventional canons) in their belief in God (1925). Is the problem really that political or moral values might be connected with a social scientist’s work or is the problem how they are connected? Further thought will show that the goal of removing either set of values from inquiry entirely is not so advisable. Many of the relevant “scientific” values are also “moral” or “ethical” values, such as fairness in judgment and honesty in presentation. Political values might also be scientific values, such as the imperative to respect the rights of a subject/citizen. None of this means that these values cannot be separated. It does mean that we should bring these values into dialog with each other when we are “doing” science.

Thus, the notion of “scientific irrelevance” is far more interesting and problematic than it might first appear. We can imagine many values that could be defined as irrelevant if we define relevance as having to do with values that designate a project as science. It is not a necessary condition of science to want to make the world a better place or to consider the discovery of the complexities of the physical and social world a sanctification of its creator. But that these values are not necessary to good science does not make them irrelevant. They are certainly not irrelevant to good scholarship.

A third issue underlying the values-science connection is the question of where in a research program values enter. This reflects the distinction often made between the “context of discovery” and the “context of justification.” There are parts of the research process where a wide range of values may fruitfully play substantial and explicit roles; and there are parts in which the values embedded in dominant scientific approaches are likely to be used by many of us to suppress other values. Tetlock does not suggest, for example, that basic research questions should be framed in ignorance of our moral or political values. Most of us would not study racism, war, cognitive complexity, gender discrimination, or even politics in general if they merely suggested unsolved empirical questions.

However, Tetlock’s specific example, and solutions, focus on the context of justification, and thus on methodological fine-tuning (in contrast to the broader issues about the nature of science and the scientific community which we have tried to raise above). It is not clear to us whether his general complaint about the influence of “good intentions” is also about justification, discovery, or merely the communication of results. Regarding the latter, one of us has experienced the problem that, regardless of what one has actually done in an inquiry, the discussion of certain topics (like “women”), or the use of particular terms and concepts,
is perceived as if one wore a banner proclaiming that one's values are excessively intrusive.

IV. THE STEPS TO GRACE

Tetlock suggests four techniques to guard against overly intrusive values. The first, "rigorous consistency in evaluating counterfactual claims," is correct, but it is already a common scholarly (not just scientific) standard. Certainly we have all seen examples of those who live up to this criterion. How far does this get us? Is rigorous consistency in evaluating counterfactuals a straightforward process? We doubt it.

Tetlock's second prescription narrows our focus to specific research designs. Embedding experiments in public opinion surveys is a splendid idea — where the question at hand is best answered in that way and if one has sufficient funding to do so. Tetlock, Sniderman, and their colleagues have obtained very useful and interesting results by using this method. He also contends that indicators should be as clean as possible. We agree, as again would most methods textbooks, although there is not likely always to be agreement on what the best or cleanest indicator is.

The advice to quantify the qualitative is more contentious. Tetlock retraces the familiar pitfalls of case studies and the potential bias and nonsystematicity of qualitative work. We agree that actual quantitative work is preferable to pseudo-quantitative research. There is plenty of "qualitative" work that uses quantitative language, discussing frequency, and distribution, and coincidence without being moderately precise about the quantitative evidence. Research indicates that people are generally not very skilled at such types of estimation, and that our cognitive acumen is easily thrown off. Quantifying the pseudoquantitative or avoiding the pretense of quantitative logic altogether is probably a good idea.

But we cannot endorse the idea that this means that it is always better to attempt quantification. Certainly, as Tetlock says, "at their worst, case studies can be unfocused, loosely reasoned, vague, impressionistic, and self-serving." But poor quantitative research is mindless, trivial, mechanical, incomprehensible, boring — as well as unfocused, loosely reasoned, vague, and the rest. Actually, quantitative methods merely add another layer of interpretation to the research process (Kritzer 1990). It is only the fact that most of our community shares a common socialization to understand r-squares or chi squares for the psychologists) that allows us to paper over this additional interpretation. It is true that problems such as generalization, reliability, and detecting spuriousness are especially difficult in qualitative methodology and that in the past, outside of anthropology and clinical psychology, qualitative method was simply an absence of numbers. The latter is no longer as widely the case. Above all, we believe a substantial amount of the research in political psychology would profit by more interpretive and multiple-layered strategies that are often likely to be qualitative in technique.

If the ultimate object of study in political psychology research is political thinking and action in the context of political institutions or collective political processes, we cannot forgo analyses incorporating holistic, contextual, and intersubjective modes of thinking. The sum of the parts we examine in fragmented modes rarely equals the whole; and it is very difficult to move between compartmentalized thinking to the whole. It is necessary to develop an analytical strategy to do so, which means interconnecting qualitative and quantitative strategies. We worry about research in which the little puzzle pieces (Popper) become the whole which, in our opinion, is a chief drawback of experimental psychology. At its best, political psychology research is probably not exclusively one or the other.

Tetlock's final advice, to search for counterexamples to the received wisdom, is noncontroversial as well as crucial not just to good science but also to good scholarship and even to good aesthetics. No one could be more vehement about the need to test the readily accepted than women's studies scholars, of whom one of us is one. There is nothing like the received wisdom on women's intelligence and cognitive skills, their personalities, sexuality, or political predilections to make one suspicious of received wisdom in science or — judging by the philosophical canon — in the humanities. The search for counterexamples is an excellent strategy in the more interpretive modes of scholarship as well, as the senior author found in working on her recent book on the political theory of Mary Wollstonecraft. In making claims about interpretations of Wollstonecraft's texts, she carefully searched them for passages that seemed to violate the interpretation.

Tetlock's prescriptions, then, offer limited means for ensuring that when specific scholars accept the mantle of science, thus claiming an openness to unanticipated and disagreeable outcomes, they actually act reasonably closely to this claim. This returns to a crucial issue of epistemology, methodology, and ethics. In asking who decides what constitutes undue value influence, we run into the problem of enforcement — by what means and by whom? We ask scientists to be honest; but we ask this of politicians and CEOs as well. Scientists, in contrast with most other citizens, demand the right to be completely self-regulating. This places an unusually high premium on professional socialization. Here lies the uncovered cornerstone of Tetlock's suggestions. In fact, the socialization of scientists to such ethical ideals is probably more thorough than that of

---

11Useful as the embedded experiment is, it is still susceptible to the "law of the hammer," that is, give a small child a hammer and very soon he or she decides everything looks like a nail.

12Of course, a "case study" is not necessarily qualitative or quantitative (Ragin, 1991). Our colleague Bert Kritzer is fond of pointing out that The American Voter is basically a case study of the 1952 and 1956 elections.
CEOs or politicians, and most conduct themselves admirably and honorably. But we are human, and to ask any human to be completely self-regulating is a weak cornerstone upon which to build a community.

V. ON THE RELEVANCE OF POLITICS AND MORALITY TO SCIENCE

Let us return to the original question, of the relationship between moral and political values. The center of gravity in our disagreement with Tetlock turns on his claim that

I take an integratively simple stand on this question. It is generally a bad idea to dilute the goal of doing high-quality science with a host of additional moral and political goals. I believe that if we fail to institutionalize checks on the overt politicization of political psychological knowledge, we sacrifice our collective credibility as a science.

Throughout Tetlock’s article, politics is at best extraneous and at worst dangerous to political psychology. Science is the core goal and good value in his story. Politics and the polity appear shadowy, at best as applications of science.

Here is perhaps where the interdisciplinary nature of political psychology shows. Our training as political scientists links us to a tradition of political theorists and philosophers who tried to understand the political world as it exists, envision better alternatives, and figure out how to get there. The moral-ethical, political, and analytic-scientific aspects of such a project remain distinct, but not entirely separable.

Political (and social) science developed through a passion for understanding how we could better govern ourselves and how we could mitigate the excesses of the small elite and the mass collective. Our eighteenth-century predecessors who engaged in what they variously defined as moral, social, and political science fostered the idea that, to know what could be, we had to know what was, and this required precision of thinking and systematic clarity of observation removed, as much as possible, from intrusive “enthusiasms” (Halévy, 1955; Farr, 1988), so long as the return to the whole, to politics, was attempted. Likewise, nearly a century ago people with a political agenda urged a science of politics in the cause of Progressivism (Ross, 1991). But the growth of scientism in social science led the community to suppress those connections. The professionalization of scholarship has separated training in normative political philosophy and empirical theory and method. Today, political psychologists typically only mention a few famous names and catch phrases to add a patina of philosophy. Normative theorizing is left to others, and reintegrating the whole is rare. This is more dangerous than the intrusion of moral or political values.

Tetlock agrees that few political psychologists are motivated purely by “epistemic” or “partisan” values. But our point is wider and deeper. Political values (partisan values being only a part of those) are not unfortunate characteristics that must be expunged from the scholarly world of political psychologists. As Tetlock argues, we can and should employ a number of techniques in the phase of discovery to make sure that we don’t find only what we want to find. When we are writing as scholars, it is also important to make sure we don’t mislead for political purposes. But as debates over expert witnesses indicate (Fiske et al., 1991), this is not simple.

Tetlock distinguishes the individual qua scientist and qua citizen (i.e., qua political and moral being). Most of us probably accept this distinction. We say things in a nonprofessional capacity that we would not say in our role as scientists. But the distinction is neither clear nor easy. It is even harder if one believes in a republican citizenship model of democratic politics, in which “the political” is not so clearly severed from other human activity.

VI. SCIENTIFIC CATHOLICISM AND SCHOLARLY ECUMENISM

We have left for the end a point that underscores the whole. Tetlock writes as if the definitions of science and epistemology were nonproblematic. They are not, even among those who call themselves scientists and prize what they understand as scientific values. Statistical and empirical research methodology is widely taught as though the statistical and epistemological theories on which they are built are self-evidently true and noncontroversial. There is a lively literature on the nature of science that should be an integral part of political psychology training. Methodology is not a canonical truth, and thoughtful devotion is not heresy. Science is not a totally separate religion from at least some other branches of good scholarship, although there are clear denominational differences.

Indeed, science is built on shared values, as are religion, politics, and morality. Scientists must find a way to cope with their existence as political and moral beings. Tetlock suggests a way that might be termed Kantian. While admitting that we can never eliminate the role of scientifically “irrelevant” values in research, he claims that we can minimize their impact (thereby saving the community of political psychology from banishment to hell) by acting as if we were politically neutral regarding our results. While admitting the impossibility of such neutrality, we move our community closer to it by acting as we would like all other scientists to act. But this solution, especially taken alone, denies the relevance of moral and political values to science, rather than engaging and using them. We close with a modest, perhaps more Hegelian proposal. The values that we possess should be brought into a meaningful relationship with one another. We recognize the full relevance of good scientific intentions only when they are juxtaposed with nonscientific ones. Hell is what we make of it; but so is heaven. Tetlock’s scientific heaven is one where irrelevant values are prevented from
interfering with the conduct of science (even if we cannot be fully saved, it is the quest that makes us scientific). Our heaven is one where moral, political, and scientific values are able to merge into something more than the sum of the parts, to help to create and transcend the limitations of themselves alone. This, unlike Tetlock’s heaven, is a goal that can be partially attained. Partial value neutrality is not value neutrality at all; but a partial merging and transcending is possible. It is the latter that we seek, as scientists and as moral and political beings.

AUTHORS’ NOTE:

Authorship is alphabetical. We both wish to acknowledge the impact of stimulating discussions in the Logic of Inquiry Study Group (1992) and the graduate seminar on Normative Political Philosophy and Empirical Political Psychology (1993) on our thinking.

REFERENCES


How Politicized is Political Psychology and Is There Anything We Should Do About It?

Philip E. Tetlock

University of California, Berkeley

This exchange pivots around two critical issues: (1) How politicized is political psychology? (2) What, if anything, should we do about it? I propose the turnabout test as a formal method for gauging the politicization of research programs (a test that can be usefully supplemented by a second procedure, the falsifiability test). Both White’s and Sears’s research programs do not fare well by these standards. I also warn of the consequences of adopting the alternative epistemologies sketched by Kroeger and Sapiro and by Sears, both of which suggest I am unduly concerned by the dangers of politicization. I conclude by sketching my own epistemological views which include a Weberian emphasis on the separation between politics and scholarship, a Popperian/Lakatosian emphasis on testable theories, and a Mehlhian stress on construct validation. From this perspective, although political psychologists can never achieve value neutrality, they can and should do a much better job than they have in monitoring and containing the impact of extraneous moral-political values on the conduct and interpretation of research.

KEY WORDS: value neutrality, symbolic racism, deterrence, political correctness, political biases in hypothesis testing.

Science is supposed to be a collective enterprise that encourages individual investigators to be more thoughtful, rigorous, and self-critical than they would have been if they had been left to their own devices. In this spirit, the current exchange of views was useful. It prompted me to be more explicit about my philosophical premises and to try to anticipate the variety of objections that reasonable critics might raise to my views. To be specific, I foresaw the following possible objections, several of which did emerge in one or another form:

A. Tetlock is wrong in claiming that specific research programs have been politicized. The work he cites satisfies the standards of good, even excellent, social science;
Tetlock is right about particular research programs but exaggerates the pervasiveness of the problem for the field in general. No one takes this position—which may be fortunate because it raises complex conceptual and sampling issues far beyond this symposium;

C. Tetlock may or may not be right that large portions of political psychology have been politicized but wrong to break the rules of scholarly civility by pointing it out;

D. Tetlock is silly to fret about the politicization of this or that research program. He does not appreciate the logical implications of accepting that a value-free social science is impossible. Political psychology always was and always will be a continuation of the political struggle through alternative (scientific) means. White, Sears, and, for that matter, Tetlock are all political animals; each trying to advance particular moral-political causes and using the rhetoric and methods of science for that purpose. This objection is the epistemological equivalent of “You really noticed! Wake up and smell the coffee”;

E. Tetlock is attempting to revive a neopositivist philosophy of science that will straitjacket scholarship into mindless empiricism and number crunching;

F. Tetlock’s methodological and conceptual prescriptions are trite and commonsensical (the stuff of basic textbooks in research methods);

G. Tetlock is just unrealistic: he wants political psychologists to be politically celibate.

In addition, three objections surfaced that, I must confess, I did not anticipate. Paul Sniderman argues that I have been too quick to assume that political motives underlie scientific blunders when a more parsimonious explanation is methodological or conceptual incompetence. He also warns against motive-mongering: Who cares about researchers’ motives? What matters is what we do, and science owes a debt even to those who make mistakes, perhaps especially big mistakes (“howlers”). Finally, Lloyd Etcherege maintains that political psychology suffers less from a surfeit of political activism than from a shortage of political courage. Etcherege takes the “scientific establishment” to task for its failure to rise to the challenge of assessing the empirical propositions underlying “Reaganomics.”

Have I been too harsh in appraising the scientific merits of certain research programs? It is inappropriate here to respond point by point to Sears’s spirited defense. As Kroeger and Sapiro note, the Sniderman-Tetlock dispute with Sears-Kinder is already old news. In my target article, I did not want to itemize, once again, the empirical, logical, and methodological objections that Sniderman and Tetlock (1986) raised (see also Schuman, Steeh, & Bobo, 1985). Sears is right, therefore, in his complaint that I selectively simplified the large literature on symbolic racism. For the record, however, I have not retracted my earlier, more comprehensive critique. It still strikes me as wrong to claim to have discovered a new racism when one fails to control statistically for the influence of old-fashioned bigotry. It still strikes me as misleading to conceive of self-interest so narrowly that it is confined to immediate threats to one’s personal well-being. Finally, it strikes me as odd in a major research program for measures of the principal variable to have almost no construct validation history—no history because the measures change from one study to the next, and the authors do not call the changes to the attention of readers. Sniderman is right that mistakes of this magnitude count as “howlers.”

My goal in the target article was, however, to zero in on the trademark feature of the research program most relevant to this exchange: namely, the deliberate blurring (construct obfuscation) of the distinction between traditional American values and racism at the item level, at the stage of scale construction (cf. Kinder, 1986). This intellectual move gives symbolic racism theory political teeth; it amounts to the assertion that racism is built into the American ethos itself (as American as apple pie). Sears is entitled to advance any hypothesis he wants, but to label a person or viewpoint as racist is a serious charge, and Sears should explain why it is legitimate to brand people as racists merely because they object to affirmative action. A major goal of political psychological analysis is to establish the extent to which opposition to affirmative action is driven by raw prejudice, and the whole project is hopeless if the former is simply defined as the latter.

Sears does not respond directly to the parable of the Symbolic Marxism scale in which conservative investigators label liberals who support the civil liberties of American communists “Symbolic Marxists.” He responds only indirectly. He notes that “in ten empirical studies, conducted over more than two decades of research, we have most often found independent effects of both racism and conservatism on whites’ opposition to racial policies, although racism generally has had stronger effects.” However, in multiple regression studies of the sort that Sears cites, much depends on exactly how one defines the conceptual and operational boundaries of constructs. I maintain that Sears and colleagues operationally define symbolic racism in ways that make it difficult for principled nonracist conservatives not to receive high scores. Sears maintains that things could not possibly be that bad because symbolic racism researchers have frequently found independent statistical effects for both symbolic racism and conservatism on whites’ opposition to racial policies. But conservatism is a complex assessment target that scholars have taken to include a loosely correlated amalgam of factual beliefs (America is already a meritocratic society; America has achieved equality of opportunity; government intervention typically
creates more problems than it solves . . . ), specific policy prescriptions (government taxes and regulations are already excessive; market mechanisms provide the most efficient solutions to societal problems), and abstract values (the importance of self-reliance, entrepreneurial initiative, hard work, obeying the law, sexual restraint, keeping families intact, and respect for property rights)—a long list of considerations that may or may not be merely code words for “symbolic racism.”

There is no logical or statistical reason why both politically tendentious symbolic racism scales and standard measures of conservatism could not consistently receive substantial regression coefficients in studies of racial policy reasoning. Moreover, the data Sears cites—the tendency for symbolic racism to have stronger effects than conservatism—may simply reflect the expropriation of causal variance due to various components of conservatism by operationally defining symbolic racism in a way that subsumes those components.

Deliberately mixing racial and ideological issues at the item level not only politicizes scholarship; it is a sure-fire method of preventing us from achieving a deeper causal understanding of American politics. A better way to proceed—a method Sears disparages as hopelessly artificial—is to embed experiments in representative sample surveys that disentangle the relative roles of race and threats to traditional or conservative values as determinants of opposition to particular policies. Experimentation, in concert with careful conceptual and empirical development of dispositional measures, promises to shed more light on the dynamics of racial politics than the confitd a priori pronouncements of partisans of the left or right. Moreover, the proof is in the pudding. The Sniderman et al. (1991) results are embarrassing both to symbolic racism theory and to those who would argue that contemporary conservatism and racism are completely unconnected.

Sears’s response (or lack thereof) to the Symbolic Marxism parable is puzzling from my epistemological position, which stresses the dangers of politicized research, but less puzzling in light of the epistemological position he stakes out for himself. In his view, social scientists should model themselves after lawyers and “take strong positions promoting one view and present evidence from that perspective.” I turn to this philosophical disagreement shortly.

Have I broken norms of scholarly civility? Although Kroeger and Sapiro decline the difficult task of evaluating the claim that partisan political preferences trumped normal scientific standards of evidence and proof in specific research programs, they do offer a lecture on manners, “especially when the criticisms entail attributions of motive and causation for which there is insufficient direct evidence or consideration of alternative hypotheses” (p. 2). Accordingly, it is instructive to examine once again the evidence and the alternative hypotheses. What are the possibilities? (1) the work I examine does not bend normal scientific standards of evidence and proof in favor of politically preferred hypotheses; (2) standards are bent, but not because of political motives—rather, the researchers just made mistakes; (3) the charges I made are essentially accurate.

With respect to the first possibility, my target article provides ample grounds for concluding that scientific standards of evidence and proof were bent in a particular political direction. The critical argument is the turnabout test. If one believes that the scale construction and validation procedures of symbolic racism researchers or White’s analysis of historical evidence satisfy normal scientific standards of evidence and proof, then one relinquishes scientific grounds for complaining when Symbolic Marxism or Peace-Through-Strength researchers use logically identical procedures to generate support for their preferred conclusions. If political psychologists are silent in the former cases (only a few protests have been registered) but apoplectic in the latter cases (virtually everyone I have talked to is indignant about the Symbolic Marxism scale), then we have conclusive evidence of a directional political bias.

With respect to the second possibility, I have too much respect for the other accomplishments of the investigators involved to suggest that they are acquainted with basic principles of scientific methodology and logical inference. Although Sniderman (1993) is unwilling to rule it out, incompetence does not strike me as a plausible hypothesis. From working with Ralph White on a chapter he contributed to Suedfeld and Tetlock (1991), I know that he knew both that there were counterinterpretations of the historical cases he advanced in support of conflict spiral theory and that there were historical cases he rarely mentioned that may support deterrence theory. As for symbolic racism researchers, it is impossible after the Sniderman and Tetlock (1986)/Kinder (1986) exchange that they were unaware that defining symbolic racism in terms of opposition to busing and affirmative action was not a political act that many would find offensive. Indeed, Kroeger and Sapiro might consider whether defining millions of conservatives as racists by operational fiat constitutes a breach of “scholarly civility.”

Where does this leave us? We can rule out that no bending of standards of evidence and proof occurred. We can rule out naivete or incompetence. We can establish that the liberal partisan preferences/schemata of the investigators were at least a necessary condition for standards to be bent in a particular direction. No one contests the factual proposition that the investigators in question (like most of us) were “liberals” (in the late-20th-century meaning of the term) or the counterfactual propositions that conservative/libertarian investigators would never have employed counterfactuals as Ralph White did or would never have operationally defined symbolic racism as Sears did. From the perspective of my argument, that is all we need to know. It does not matter one whit whether the bending of standards was intentional (the researchers self-consciously decided to do some work to help out the NAACP or Physicians for Social Responsibility) or unintentional (the researchers were so convinced that Reagan’s policies were increasing the likelihood of nuclear war that American racism had mutated
into a new, more subtle, and insidious form that they thought they were simply describing the world as it was). Indeed, it does not even matter for epistemological purposes whether White and Sears are ultimately proven right in their causal claims. What matters, as Sniderman notes, is that our peer review system-our collective quality control system—failed us. That failure threatens our viability as an autonomous, self-correcting science.

How have I failed to see the logical implications of my own admission that value-free social science is impossible? Virtually all political psychologists agree that our moral and political assumptions can shape how we ask and answer research questions. Disagreements revolve around what, if anything, we should do about it. In another recent case study of politicized scholarship, Robert Packenham (1992) notes that we have essentially two options. “The first is to relax and enjoy it; to accept and encourage the fusion of scholarship and politics. The second is to try to optimize scholarly values in the conventional sense and minimize the distorting effects of political partisanship.” The first position captures the spirit of Kroeger and Sapiro’s recommendations for our discipline. In their view, I chose the wrong fork of the road when I urged trying to approximate value neutrality by more rigorously enforcing standards of evidence and proof and by sharply distinguishing between what we speak as scientists and as citizens. They offer an alternative vision:

A. “Our heaven is one where moral, political, and scientific values are able to merge into something more than the sum of the parts, to help create and transcend the limitations of themselves alone”;

B. “The values that we possess should be brought into meaningful relationship with each other”;

C. We should bring scientific and political values into “dialogue” with each other when we are doing science.

I find the vagueness of key words in this argument—such as “merge,” “transcend,” “meaningful relationship,” and “dialogue”—worrysome. Kroeger and Sapiro, in effect, offer an open-ended invitation to do science and politics simultaneously. This “merger” recipe contrasts sharply with my warnings concerning diluting scientific goals with political ones. I leave it to readers to imagine how politicized political psychology would have to become before the merger of science and politics became a hostile takeover. Would conservatives who constructed a Symbolic Marxism scale to predict egalitarian policy preferences count as an example of a fruitful “dialogue” between scientific and extra-scientific values? Would expanding the concept of rape to encompass most heterosexual relationships count as a constructive engagement of political and knowledge-seeking goals? Is there anything special about science—anything that cannot be merged, compromised, or subordinated to other goals?

Political psychology can make for odd epistemological bedfellows. Where-

as Kroeger and Sapiro take me to task for arguing that our collective credibility as a science depends on self-critical efforts to monitor and minimize the influence of extraneous political values on inquiry (a mistake that puts me on my own personal road to hell), Sears agrees “wholeheartedly” with that part of my argument (illustrating, alas, how easily one can be doubly damned in exchanges of this sort).

Sears and I soon diverge, however, on the question of how “weak constructivists” should go about enforcing scientific norms of evidence and procedure. He suspects I prefer “impartial judges” who rise above the heat of the theoretical and political struggle to announce “dispassionate decisions.” He prefers an adversarial system in which “protagonists take strong positions promoting one view, and present evidence from that perspective. . . . Over time, it is felt, the truth will out, and better so for having the best possible advocates passionately arguing each position.”

My view is actually that both the inquisitorial and adversarial systems have drawbacks. Advocates of the inquisitorial system must answer the painfully obvious question: Who qualifies as an impartial judge? Advocates of the adversarial system must explain how they can achieve reasonable balance when the distribution of political opinion is as skewed as that among political psychologists. I doubt very much that White’s work on war and peace or Sears’s work on symbolic racism would have appeared in the forms they did if these investigators had had to contend with a conservative as well as liberal reviewers. Accountability to a more politically diverse collection of reviewers would have checked the intrusion of blatant political bias. But where do we find the conservatives? There is also another deeper difficulty with the adversarial proposal. Do we want American attorneys of the late 20th century to become our epistemological role models? Is it acceptable to be as mercenary, as ideological, and as client-driven as many trial lawyers? Do we not want scientists to internalize some standards of evidence and procedure? Am I merely promoting pretentious “scientism” (white-lab-coated hypocrisy) when I encourage both individual investigators and the research community to aspire to objectivity and value-neutrality?

I do not claim to have identified the optimal mixture of inquisitorial and adversarial accountability for preserving political psychology as a reasonably autonomous, self-correcting scientific enterprise. I do not even claim that the current system has collapsed (this symposium is testimony to that). I do claim that the current system shows systematic signs of political favoritism.

Are we up to policing ourselves? Kroeger and Sapiro believe that they have discovered the uncovered cornerstone of my argument: namely, that scientists are up to the task of policing themselves. They note that “we are human and to ask any human—or group of humans—to be “completely self-regulating” is a weak cornerstone upon which to build a community.” This remark suggests that I have not been as clear as I should have been. First, my original article asserts that we
political psychologists have been either unwilling or unable to prevent the intrusion of extrascientific values into our scientific work. My faith in "professional socialization" is therefore not unconditional. Second, I do not believe that individual investigators are up to the task of policing themselves. The elaborate system of peer review for research grants and journal publications is a necessary quality-control mechanism (although it can exclude thoughtful dissenters as well as crackpots). Third, I fear that effective scientific policing may well fail even with the very best of falsificationist intentions. Although we political psychologists like to trumpet the virtues of diversity—gender, ethnic, and intellectual—we are in key respects quite homogeneous. There are few defenders of neoclassical economics or of rationality as a good first-order approximation of human nature. There are few defenders of evolutionary psychology or sociobiology. There are few advocates of neorealism or of deterrence theory. I am not convinced that certain points of view can receive a fair hearing in our discipline. Count me, therefore, as firmly agnostic on the question of self-policing. Much will hinge on our willingness to ask ourselves and each other morally and scientifically dissonant questions that few seem, at present, disposed to pose.

Is my methodological advice commonsensical, even trite? I did not intend the research procedures I proposed for guarding against excessive intrusion of values to count as an exhaustive or especially innovative list. Kroeger and Sapiro certainly see nothing innovative in it. They note that rigorous consistency in evaluating counterfactual claims is already a common scholarly standard—perhaps, but not common enough, as my commentary on White's work demonstrates. More curious, they think that my advice "to search for counterexamples to the received wisdom" is noncontroversial. They note that "no one could be more vehement about the need to test the readily accepted than women's studies scholars, of whom one of us is one" (p. 563). Kroeger and Sapiro are right that feminist scholars have usefully challenged a variety of preconceptions and shibboleths concerning women's cognitive styles and abilities, their personalities, and their sexuality. My own conception of science is, however, more dialectical than that of Kroeger and Sapiro. It does not take long for a new conventional wisdom to arise and for yesterday's revolutionaries to become today's thought police.

If I believed political psychologists were as rigorously consistent in evaluating counterfactuals and as scrupulously openminded to counterexamples as do Kroeger and Sapiro, I would not have bothered to write the target article. A core finding of cognitive psychology is how hard most people find it to answer the straightforward question: what would it take to convince you that you are wrong? History suggests that scientists are no exception to this generalization; even relatively apolitical scientists often cling to favorite theories in the face of mounting contradictory evidence. My hunch is that politicization typically makes matters worse, exacerbating the human tendency to belief perseverance. It is not surprising, from this perspective, that White and others have been conspicuously silent on how his model of international conflict—a variant of spiral theory—should be modified in response to disconfirmation of the once-popular prediction that Reagan's "confrontational" policies would exacerbate the Cold War. Equally silent have been White's counterparts on the political right: advocates of the essentialist-totalitarian view of the Soviet polity who for many years glibly dismissed Gorbachev as a "neo-Stalinist in Gucci garb." In a similar vein, Sears is quick to dismiss the experimental tests by Sniderman et al. (1991) of the symbolic racism "fusion" hypothesis, but never gets around to telling us what kinds of evidence would induce him to abandon symbolic racism theory.

Do I seek to impose a dogmatic positivism on political psychology? Kroeger and Sapiro are even less impressed by the other two methodological nostrums I advanced: embedded experiments in surveys (to disentangle otherwise hopeless controversies over causality) and, when possible, quantify case studies (to assess the amount and types of consensus among historical observers on various political psychological claims such as groupthink). They fear overreliance on experimentation (reminding us of the law of the hammer), and they attribute the ridiculous position to me that "it is always better to attempt quantification" (overlooking the fact that the Q-sort analyses of groupthink and vigilant episodes would have been impossible without a rich network of qualitative case studies to draw upon). For the record, therefore, I recognize that quantitative research can be trivial and qualitative research inspiring. I do not seek to impose a narrowly-minded, number-crunching positivism or a false dichotomy between qualitative or quantitative methods which, I have argued elsewhere, possess largely complementary strengths and weaknesses (cf. Tetlock, 1983).

Am I advocating political celibacy? Another misunderstanding of my argument is that political psychologists should abstain from policy controversies in general. Nothing could be further from the truth. The research community possesses knowledge of cause-effect relations and analytical skills that can prove valuable in a wide range of policy controversies, from affirmative action to daycare to international negotiations to taxation. The key question is: in what capacity do we enter such controversies? Do we blur the roles of scientist (whose aim is truth) and advocate (whose aim is justice)? Or, do we attempt some separation of roles? My view is that of Koshland (1990), the editor of Science: "Scientists are the servants of society, not its masters, and we should remain so" (p. 9). We serve neither science nor society well when we (accidentally or intentionally) smuggle our political values into our research findings.

Koshland's advice to separate our scientific and political goals is not just empty exhortation. Although the separation can never be perfect, many political psychologists manage to make the distinction. Robert Jervis and Alexander George certainly have political preferences, but they do not allow these preferences to bias their treatment of arguments and evidence in the flagrant fashion
that Ralph White did. Sniderman, Carmines, Kleugel, Smith, Schuman, Bobo, and others are probably no less opposed to racism than Sears and Kinder, but they refrain from stigmatizing as racists political groups with whom they disagree. It is also worth noting that our knowledge of how to distinguish factual and value judgments has grown apace over the last 30 years. For example, Hammond, Harvey, and Hastie (1992) offer exemplary illustrations of how to use signal detection theory and social judgment theory to separate experts’ judgments of the likelihood of events from policy-makers’ value judgments of the relative importance of avoiding false positive errors (taking action when we should not have) and false negative errors (not taking action when we should have).

In short, I am no advocate of political celibacy, but I am a critic of political promiscuity—the indiscriminate mingling of factual and value claims. Doing political psychology well requires a measure of political self-restraint.

CONCLUDING REMARKS

Some scholars may view this exchange as but one more manifestation of a broader political struggle in American universities of the 1980s and ‘90s: between liberal activists and conservative traditionalists. Others may lament the lapse in academic etiquette, the regrettable “personal” tone of the exchange. Both views are off the mark. First, I would be every bit as disturbed if Sears had employed the same scale construction and hypothesis testing procedures in a research program organized around the concept of symbolic Marxism or if White had selectively employed counterfactuals to argue that Reagan deserved sole responsibility for winning the Cold War. Second, we should not be too thin-skinned. I do not take personal offense when David Sears invites readers to speculate on whether my own political views may be responsible for the “selectivity” of my review of symbolic racism work. I agree with Sears that we too often pretend that we have somehow transcended politics. With the notable exception of Sniderman (1993), we all now seem to be “constructivists” of one sort or another who accept the impossibility of value neutrality and the omnipresent possibility that preconceptions may shape the design, conduct, and interpretation of our research. There is no reason to take offense when someone poses the obvious and perfectly fair question: are you not pursuing a political agenda through your scholarly research?

My answer to that question is “not to the best of my knowledge.” I looked for but could not find influential research programs in political psychology that bore the imprint of conservative views. Perhaps some more observant scholar will soon correct me. So be it. It is quite reasonable to scrutinize each other’s work for possible normative biases: to document that “if investigator X had focused on this rather than that case, or applied this procedure more consistently,

or given more weight to this logical or empirical counterargument, or acknowledged this design flaw, she would not have made this error—which, incidentally, neatly reinforces her worldview.” Of course, such exercises in collective self-criticism can become trite: we ritualistically agree that we are all guilty of overt or covert partisanship and then all agree to absolve ourselves. Or the exercise can become malicious. We set out on neo-McCarthyite vendettas to destroy each other’s careers. In either case, I will have done the profession a disservice (the outcome Sniderman fears). If, however, the target article alerts us to our collective blind spots and promotes more rigorously self-critical and even-handed enforcement of scientific norms, I count that as success.

REFERENCES


