

# DEMOCRACY FOR REALISTS

*Why Elections Do Not Produce  
Responsive Government*

*With a new afterword by the authors*

CHRISTOPHER H. ACHEN  
LARRY M. BARTELS

PRINCETON UNIVERSITY PRESS  
PRINCETON AND OXFORD



PRINCETON STUDIES IN  
**Political Behavior**  
*Edited by Tali Mendelberg*

the theory of social choice more generally—has no curb appeal at all: even a perfectly rational, highly informed median voter does not get what she wants. That result raised fundamental logical problems for the populist ideal by calling into question how *any* sort of electoral process could reliably aggregate potentially complex individual preferences into a coherent “will of the people.”

#### PUBLIC OPINION AND POLITICAL IDEOLOGY

Zaller's allusion to the “rationally ignorant median voter” fused two distinct aspects of Downs's “economic theory of democracy.” One is the unidimensional spatial model of electoral competition, in which parties have strong incentives to converge on the ideological “ideal point” of the median voter. The other is Downs's analysis of political information costs, which led him to conclude that, because of “the infinitesimal role which each citizen's vote plays in deciding the election,” the returns to acquiring political information “are so low that many rational voters [will] refrain from purchasing any political information *per se*” (Downs 1957, 258). Thus, “A large percentage of citizens—including voters—do not become informed to any significant degree on the issues involved in elections, even if they believe the outcomes to be important” (Downs 1957, 298).

Unfortunately, for the spatial model of electoral competition to work, “rationally ignorant” voters do need *some* political information. In particular, if they are to succeed in voting for the party closest to them they need to know their own preferences and the platforms of the competing parties regarding “the issues involved in elections.” The voters' own preferences, especially, are often simply taken for granted in the populist theory of democracy. But what if voters don't really know what they want? In that case, the folk theory of democracy, and the spatial model in particular, loses its starting point.

One telling indication that this foundation of the folk theory may be shakier than it appears is the fact that expressed political attitudes can be remarkably sensitive to seemingly innocuous variations in question wording or context. For example, 63% to 65% of Americans in the mid-1980s said that the federal government was spending too little on “assistance to the poor”; but only 20% to 25% said that it was spending too little on “welfare” (Rasinski 1989, 391). “Welfare” clearly had deeply negative connotations for many Americans, probably because it stimulated rather different mental images than “assistance to the poor” (Gilens 1999). Would additional federal spending in

this domain have reflected the will of the majority, or not? We can suggest no sensible way to answer that question.

It seems tendentious to insist that “welfare” and “assistance to the poor” denoted *different* policies, and that Americans carefully opposed the former while supporting the latter. However, even if that distinction is accepted, qualitatively similar framing effects appear in cases where the substantive distinction between alternative frames is even more tenuous. For example, in three separate experiments conducted in the mid-1970s, almost half of Americans said they would “not allow” a communist to give a speech, while only about one-fourth said they would “forbid” him or her from doing so (Schuman and Presser 1981, 277). In the weeks leading up to the 1991 Gulf War, almost two-thirds of Americans were willing to “use military force,” but fewer than half were willing to “engage in combat,” and fewer than 30% were willing to “go to war” (Mueller 1994, 30). Framing more abstract quantitative choices in different but mathematically equivalent ways also produces predictable—and sometimes dramatic—differences in results (Pruitt 1967; Tversky and Kahneman 1981).

The psychological indeterminacy of preferences revealed by these “framing effects” (Kahneman, Slovic, and Tversky 1982) and question-wording experiments calls into question the most fundamental assumption of populist democratic theory—that citizens have definite preferences to be elicited and aggregated through some well-specified process of collective choice (Bartels 2003). In this respect, modern cognitive psychology has sharpened and reinforced concerns about the quality of public opinion raised by critics of democracy from Plato to the pioneering survey researchers of the 1940s and 1950s.

The first rigorous scientific portrait of the American voter, by Bernard Berelson and his colleagues at Columbia University, found that “the voter falls short” of displaying the motivation, knowledge, and rationality expected by “traditional normative theory” (Berelson, Lazarsfeld, and McPhee 1954, 308, 306). “On the issues of the campaign,” the Columbia scholars found (Berelson, Lazarsfeld, and McPhee 1954, 309, 311), “there is a considerable amount of ‘don't know’—sometimes reflecting genuine indecision, more often meaning ‘don't care.’” Voters consistently misperceived where candidates stood on the important issues of the day and exaggerated the extent of public support for their favorite candidates. And vote choices were “relatively invulnerable to direct argumentation” and “characterized more by faith than by conviction and by wishful expectation rather than careful prediction of consequences.”

Several years later, in a landmark study of *The American Voter*, Angus Campbell and his colleagues at the University of Michigan described "the general impoverishment of political thought in a large proportion of the electorate." They acknowledged that "many people know the existence of few if any of the major issues of policy," much less how the competing parties and candidates might address them (Campbell et al. 1960, 543, 168, 170). Shifts in election outcomes, they concluded, were largely attributable to defections from long-standing partisan loyalties by relatively unsophisticated voters with little grasp of issues or ideology.

Philip Converse's (1964) essay on "The Nature of Belief Systems in Mass Publics" provided an even more devastating and influential portrait of the political thinking of ordinary citizens.<sup>9</sup> Employing the growing store of data collected by the Michigan Survey Research Center, Converse concluded that many citizens "do not have meaningful beliefs, even on issues that have formed the basis for intense political controversy among elites for substantial periods of time" (Converse 1964, 245).

Converse's evidence was of three kinds. First, he scrutinized respondents' answers to open-ended questions about political parties and candidates for evidence that they understood and spontaneously employed the ideological concepts at the core of elite political discourse. He found that about 3% of voters were clearly classifiable as "ideologues," with another 12% qualifying as "near-ideologues"; the vast majority of voters (and an even larger proportion of nonvoters) seemed to think about parties and candidates in terms of group interests or the "nature of the times," or in ways that conveyed "no shred of policy significance whatever" (Converse 1964, 217–218; also Campbell et al. 1960, chap. 10).<sup>10</sup>

Second, Converse assessed the degree of organization of political belief systems, as measured by statistical correlations between responses to related policy questions. Could respondents give consistently liberal or consistently conservative responses? He found only modest correlations (averaging just .23 among domestic policy views (regarding employment, aid to education, and federal housing), similarly modest correlations among foreign policy views

9 Kinder and Kalmoe (n.d., chaps 1–2) noted that "The Nature of Belief Systems in Mass Publics" had been cited almost 700 times in 2013 alone, its 50th year in print.

10 We see this classification of citizens in less hierarchical terms than Converse did. As we show in chapter 10, the well-informed people who disproportionately occupy the top rungs of his scale are "ideologues" in some of the unfortunate senses of that term as well.

(regarding foreign economic aid, soldiers abroad, and isolationism), and virtually no correlation between views across these two domains. Nor were specific policy views strongly correlated with party preferences (averaging just .07). In each case, the corresponding correlations were much higher for a sample of congressional candidates responding to related but more specific policy questions. Converse (1964, 228) interpreted these results as providing strong support for the hypothesis that "constraint among political idea-elements begins to lose its range very rapidly once we move from the most sophisticated few toward the 'grass roots.'"

Converse himself recognized that "constraint among political idea-elements"—especially across issue domains—was primarily a matter of social learning rather than logical reasoning. Critics at the time and since have pointed out that there may be nothing particularly sophisticated about parroting the specific combination of issue positions defining a conventional ideology or party line. Perhaps ordinary citizens' issue preferences lacked "constraint" because they had thoughtfully constructed their own personal political belief systems transcending conventional ideologies and party lines?

Alas, this argument ran aground on Converse's third set of analyses, which assessed the *stability* of their attitudes regarding specific issues. Converse gauged "the stability of belief elements" by tracking the same people's responses to the same questions across three separate interviews conducted at two-year intervals between 1956 and 1960. Successive responses to the same questions turned out to be remarkably *inconsistent*. The correlation coefficients measuring the temporal stability of responses for any given issue from one interview to the next ranged from a bit less than .50 down to a bit less than .30, suggesting that issue views are "extremely labile for individuals over time" (Converse 1964, 240–241).<sup>11</sup> In marked contrast, expressions of party

11 Some of this temporal instability no doubt reflects measurement error due to the inevitable vagueness of survey questions (Achen 1975). Moreover, people may bring different relevant considerations to bear in answering the same question in successive interviews, producing unstable responses even as their underlying stores of relevant considerations remain unchanged (Zaller 1992). In the first years after Converse wrote, the responses of sophisticated people (as measured by formal education) seemed to be almost as unstable as those of less sophisticated people, suggesting that the inevitable noise in survey questions was at fault. However, later studies using better measures of political sophistication (based on respondents' demonstrated factual knowledge about politics) have generally found the opinions of more sophisticated people to be a good deal more stable. For example, Kinder and Kalmoe (n.d., ms. 70) calculated that the average stability of responses to five issue questions in the 1992–1996 American National Election Studies panel survey ranged from .60 among the best-informed quintile of respondents down to .25 in the bottom information quintile. Again, we

identification were much more stable, with correlations from one survey to the next exceeding .70. Converse (1964, 241) inferred that parties "are more central within the political belief systems of the mass public than are the policy ends that the parties are designed to pursue."

Converse's essay set off a very substantial debate about his methodology and interpretations. Perhaps no single argument he advanced was fully persuasive. But even most critics agreed with his conclusion: the political "belief systems" of ordinary citizens bore little resemblance to the ideal embodied in the folk theory of democracy. As Kinder and Kalmoe (n.d., ms. 13) summarized current scholarly understanding, "Genuine ideological identification—an abiding dispositional commitment to an ideological point of view—turns out to be rare. Real liberals and real conservatives are found in impressive numbers only in the higher echelons of political society, confined to the comparatively few who are deeply and seriously engaged in political life." For most ordinary citizens, ideology is—at best—a byproduct of more basic partisan and group loyalties. Thus, as Kinder and Kalmoe (n.d., ms. 12) noted, "Americans are much more resolute in their identification with party than they are in their identification with ideology."

Research in other countries has generally produced similar portraits of democratic citizens. Their ideological self-placements are often driven more by partisanship than by policy positions (Inglehart and Klingemann 1976). Indeed, left-right terms are sometimes meaningful primarily as alternate names for the political parties—often, names that the parties themselves have taught the voters (Arian and Shamir 1983). Even in France, the presumed home of ideological politics, Converse and Pierce (1986, chap. 4) found that most voters did not understand political "left" and "right." When citizens do understand the terms, they may still be uncertain or confused about where the parties stand on the left-right dimension (Butler and Stokes 1974, 323–337). Perhaps as a result, their partisan loyalties and issue preferences are often badly misaligned. In a 1968 survey in Italy, for example, 50% of those who identified with the right-wing Monarchist party took left-wing policy positions (Barnes 1971, 170).

Lest younger readers be tempted to suppose that this sort of confusion is a remnant of an older and less sophisticated political era (or an artifact of older and less sophisticated scholarly analysis), we note that careful recent studies

---

want to emphasize that it is a mistake to view people at the top of the information scale as sophisticated independent thinkers, as we demonstrate in the remainder of this book. The point is simply that they are more consistent in the positions they espouse from one survey to the next.

have repeatedly turned up similar findings. For example, Elizabeth Zechmeis-ter (2006, 162) found "striking, systematic differences . . . both within and across the countries" in the conceptions of "left" and "right" offered by elite private college students in Mexico and Argentina, while André Blais (personal communication) found half of German voters unable to place the party called "Die Linke"—the Left—on a left-right scale.<sup>12</sup>

This rather bleak portrait of public opinion has provoked a good deal of resistance among political scientists, and a variety of concerted attempts to overturn or evade the findings of the classic Columbia and Michigan studies.<sup>13</sup> In the 1970s, for example, some scholars claimed to have discovered *The Changing American Voter*, a much more issue-oriented and ideologically consistent specimen than the earlier studies had portrayed (Nie, Verba, and Petrocik 1976). Unfortunately, further scrutiny revealed that most of the apparent change could be attributed to changes in the questions voters were being asked rather than to more elevated political thinking. When 1970s voters were asked the old questions, their responses displayed little more consistency or sophistication than they had in the 1950s (Bishop, Oldendick, and Tuchfarber 1978a; 1978b; Brunk 1978; Sullivan et al. 1979).

Other scholars have argued that overarching ideological convictions are unnecessary because citizens can derive meaningful policy preferences from somewhat narrower "core values" such as equal opportunity, limited government, or traditional morality (Feldman 1988; Goren 2001). Citizens' allegiances (or antipathies) to these values do tend to be somewhat more stable than their specific policy preferences. However, they are a good deal less stable than the phrase "core values" would seem to imply, being significantly colored by party identification and even by short-term vote intentions (Goren 2005; McCann 1997).

Since the 1980s, the American political system has seen a substantial increase in partisan polarization, with Democratic elites becoming more clearly and consistently liberal and Republican elites more clearly and consistently

12. Even among professional politicians and intellectuals, the meaning and salience of ideology sometimes vary greatly with the political context. For example, the specific policy preferences of Latin American legislators typically explain less than 10% of the variance in their left-right self-placements (Zechmeister 2010, 105–110). In France, Converse and Pierce (1986, 129–132) found that "left" sounded good even to rightist deputies. Terms like "left" and "right" or "liberal" and "conservative," when they make sense at all, often turn out to represent partisan commitments as much or more than issue positions. Thus, ideological language need not have much genuine ideological content, even among elites.

13. Bartels (2010) summarized these and other developments in the scholarly study of American electoral behavior.



conservative (Poole and Rosenthal 2007; Theriault 2008; Mann and Ornstein 2012). As a result, rank-and-file partisans have increasingly come to adopt ideological labels consistent with their partisanship (Layman, Carsey, and Horowitz 2006; Hetherington 2009; Levendusky 2009). Has this solved Converse's problem? Alas, voters' policy preferences seem to have become only modestly more "ideologically coherent" as a result, with the average correlation between pairs of policy preferences increasing from a paltry .16 in 1972 to a slightly less paltry .20 in 2012. Not much has changed.<sup>14</sup>

Similarly, when a group of scholars half a century later painstakingly replicated many of the specific analyses presented in *The American Voter* using survey data from 2000 and 2004, they found no change in most respects, and only glacial improvements in the remainder (Lewis-Beck et al. 2008). Voters, it seems, are what they are, and not what idealistic proponents of popular sovereignty might wish them to be. The folk theory is of little use in understanding actual democratic politics.

Thus, most contemporary scholars of public opinion have come to accept, at least in broad outline, Converse's portrait of democratic citizens. Kinder and Kalmoe (n.d., ms. 61–62) conclude that "Converse's conclusion of ideological innocence still stands. . . . Educational transformation, party polarization, revolutionary changes in information dissemination, fundamental alterations in gender and race relations: impressive as these changes have been, equally impressive is how little visible effect they have had on how the American electorate understands politics."<sup>15</sup>

#### POLITICAL IGNORANCE, HEURISTICS, AND "THE MIRACLE OF AGGREGATION"

Confusion regarding political ideology is just the tip of a large iceberg of political unawareness. Michael Delli Carpini and Scott Keeter (1996) surveyed responses to hundreds of specific factual questions in U.S. opinion surveys over the preceding 50 years to provide an authoritative summary of

14 Kinder and Kalmoe (n.d., ms. 45–46) helpfully calculated that if "ideological constraint continues to increase into the indefinite future" at the same modest rate, "the American public's views on policy would eventually come to approximate the degree of structure shown by partisan elites today" — in about 300 years.

15 Converse's own subsequent assessments of these issues (1990; 2000; 2006) were broadly consistent with Kinder and Kalmoe's. (The last of these assessments appeared as part of an extensive symposium in *Critical Review* on "democratic competence," which also reprinted Converse's original essay.)

*What Americans Know about Politics and Why It Matters*. In 1952, Delli Carpini and Keeter found, only 44% of Americans could name at least one branch of government. In 1972, only 22% knew something about Watergate. In 1985, only 59% knew whether their own state's governor was a Democrat or a Republican. In 1986, only 49% knew which one nation in the world had used nuclear weapons (Delli Carpini and Keeter 1996, 70, 81, 74, 84). Delli Carpini and Keeter (1996, 270) concluded from these and scores of similar findings that "large numbers of American citizens are woefully underinformed and that overall levels of knowledge are modest at best." Robert Luskin (2002, 282) put the same conclusion rather more colorfully, observing that most people "know jaw-droppingly little about politics."

Here, too, it is striking how little seems to have changed in the decades since survey research began to shed systematic light on the nature of public opinion. Changes in the structure of the mass media have allowed people with an uncommon taste for public affairs to find an unprecedented quantity and variety of political news; but they have also allowed people with more typical tastes to abandon traditional newspapers and television news for round-the-clock sports, pet tricks, or pornography, producing an increase in the *variance* of political information levels but no change in the *average level* of political information (Baum and Kernell 1999; Prior 2007). Similarly, while formal education remains a strong predictor of individuals' knowledge about politics, substantial increases in American educational attainment have produced little apparent increase in overall levels of political knowledge. When Delli Carpini and Keeter (1996, 17) compared responses to scores of factual questions asked repeatedly in opinion surveys over the past half century, they found that "the public's level of political knowledge is little different today than it was fifty years ago. Given the ample reasons to expect changing levels of knowledge over the past fifty years, this finding provides the strongest evidence for the intractability of political knowledge and ignorance." Ilya Somin (2013, 192) concluded from a more recent survey along similar lines that "widespread political ignorance is a serious problem for democracy," and questioned "whether the modern electorate even comes close to meeting the requirements of democratic theory."

Some critics of this perspective have supposed that opinion surveys significantly underestimate people's political knowledge by providing insufficient motivation for them to answer questions correctly (Prior and Lupia 2008; Bullock et al. 2013). Unfortunately, insufficient motivation is endemic to mass politics, not an artifact of opinion surveys; we do not doubt that voters would be better informed if they were paid to learn political facts, but that

seems impractical (and, judging by the results of these studies, extremely expensive). Others imagine that “visual political knowledge”—recognizing the faces of political figures but not their names—provides “a different road to competence” (Prior 2014); but adding photographs to the ballot would raise significant additional problems of voter bias (Todorov et al. 2005; Lawson et al. 2010; Olivola and Todorov 2010; Lenz and Lawson 2011).

Most attempts to “redeem” the electorate have taken a different tack, acknowledging that voters are generally inattentive and uninformed but denying that the quality of their political decisions suffers much as a result. For example, formal theorists have proposed versions of the spatial model of elections in which the usual postulate that voters are fully informed is loosened somewhat. Unfortunately, even “uninformed” voters in these models know a great deal more than most real voters do. For example, one influential spatial model of elections with “uninformed” voters (McKelvey and Ordeshook 1985; 1986) posited that all voters know the distribution of voters’ ideal points, informed voters know the candidates’ positions exactly, and uninformed voters know the levels of “informed” support for candidates (from poll data) and the left-right order of the candidates’ positions, from which they then proceed to *infer* the candidates’ positions on the basis of spatial theory. Alas, as we have seen, most voters do not know what political “left” and “right” mean, much less know what informed voters think. Thus, few if any of these cheery assumptions are likely to hold in practice.

In the early 1990s, a spate of books with such reassuring titles as *The Reasoning Voter* (Popkin 1991), *Reasoning and Choice* (Sniderman, Brody, and Tedlock 1991), and *The Rational Public* (Page and Shapiro 1992) argued that voters could use “information shortcuts” or “heuristics” to make rational electoral choices even though they lacked detailed knowledge about candidates and policies. These shortcuts could take a variety of forms, including “cues” from trusted individuals or groups, inferences derived from political or social stereotypes, or generalizations from personal experience or folk wisdom.

Sociologists and political scientists have long recognized that citizens sometimes take “cues” from better informed friends, relatives, neighbors, or coworkers (Katz and Lazarsfeld 1955; Huckfeldt and Sprague 1995; Murtz 2006). The writer Calvin Trillin once had such an arrangement. “Mrs. Trillin, Alice, gave Cyprus to Mr. Trillin for his birthday; for the next 12 months, she would think about Cyprus and he wouldn’t have to. Mr. Trillin, for Christmas, gave Mrs. Trillin Iran. Neither of them was willing to take over thinking about the SALT [disarmament] talks” (Leonard 1982). The very humor of the story points to one limitation of this defense of democracy: while

Mr. and Mrs. Trillin were sufficiently well informed (and, we hope, politically compatible) to make such a division of labor feasible and efficient, most citizens don’t have a Cyprus watcher in the house and would not know what to make of advice about Cyprus if they had it.

The literature on “heuristics” in political science is an odd stepchild of the corresponding literature in psychology. Psychologists have devoted exhaustive attention to the biases in judgment produced by reliance on specific, identifiable heuristics. For example, the classic collection of essays edited by Daniel Kahneman, Paul Slovic, and Amos Tversky (1982) included reports on “belief in the law of small numbers,” “shortcomings in the attribution process,” “egocentric biases in availability and attribution,” “the illusion of control,” and “overconfidence in case-study judgments,” among other topics. It also included a series of essays on “corrective procedures” intended to mitigate the effects of these various biases and shortcomings.

Political scientists, by comparison, have typically been much more likely to view “heuristics” as a boon to democracy. We suspect, along with James Kuklinski and Paul Quirk (2000, 154), that that enthusiasm has much to do with the fact that “the notion of a competent citizenry is normatively attractive. It buttresses efforts to expand citizen participation and credits the citizenry for some of American democracy’s success.”

When students of political heuristics have defined the tasks of citizens sufficiently clearly for concrete performance benchmarks to be meaningful, they have tended to present those tasks in such highly simplified form that all of the difficulties of real political inference are abstracted away (Lupia and McCubbins 1998). More often, observed preferences and behavior are deemed “rational” simply because they look reasonable or seem to be influenced by plausibly relevant considerations. In one of the most colorful examples of a political “information shortcut,” Samuel Popkin argued that Mexican-American voters had good reason to be suspicious of President Gerald Ford when he made the mistake, during a Texas primary campaign appearance, of trying to down a tamale without first removing its cornhusk wrapper. According to Popkin (1991, 3), “Showing familiarity with a voter’s culture is an obvious and easy test of ability to relate to the problems and sensibilities of the ethnic group and to understand and care about them.” An obvious and easy test, yes. An accurate basis for inferring Ford’s sensitivities toward Mexican-Americans? We have no idea.

Lacking any objective standard for distinguishing reliable political cues from unreliable ones, some scholars have simply asked whether uninformed citizens—using whatever “information shortcuts” are available to them—

manage to mimic the preferences and choices of better informed people. Alas, statistical analyses of the impact of political information on policy preferences have produced ample evidence of substantial divergences between the preferences of relatively uninformed and better informed citizens (Delli Carpini and Keeter 1996, chap. 6; Althaus 1998). Similarly, when ordinary people are exposed to intensive political education and conversation on specific policy issues, they often change their mind (Luskin, Fishkin, and Jowell 2002; Sturgis 2003).

Parallel analyses of voting behavior have likewise found that uninformed citizens cast significantly different votes than those who were better informed. For example, Bartels (1996) estimated that actual vote choices fell about halfway between what they would have been if voters had been fully informed and what they would have been if everyone had picked candidates by flipping coins.<sup>16</sup> Richard Lau and David Redlawsk (1997; 2006) analyzed the same elections using a less demanding criterion for assessing "correct" voting. (They took each voter's partisanship, policy positions, and evaluations of candidate performance as given, setting aside the fact that these, too, may be subject to errors and biases.) They found that about 70% of voters, on average, chose the candidate who best matched their own expressed preferences. Lau and Redlawsk (2006, 88, 263) wondered, "Is 70 percent correct enough?"

Answering that question requires a careful assessment of the extent to which "incorrect" votes skew election outcomes. Optimism about the competence of democratic electorates has often been bolstered (at least among political scientists) by appeals to what Converse (1990) dubbed the "miracle of aggregation"—an idea formalized by the Marquis de Condorcet more than 200 years ago and forcefully argued with empirical evidence by Benjamin Page and Robert Shapiro (1992). Condorcet demonstrated mathematically that if several jurors make independent judgments of a suspect's guilt or innocence, a majority are quite likely to judge correctly even if every individual juror is only modestly more likely than chance to reach the correct conclusion.

16 The phrase "fully informed" is a misnomer here, since Bartels's imputations of "fully informed" voting behavior were based on observed variations in voting behavior across a five-point summary measure of survey respondents' general level of information about politics and public affairs. It seems safe to assume that even respondents at the top of this information scale were, in reality, far from being "fully informed." Thus, the effects of low political knowledge on voting behavior were almost certainly underestimated. Bartels (1990) provided a more detailed discussion of political interests, political enlightenment, and the logic and potential applications of the imputation strategy.

Applied to electoral politics, Condorcet's logic suggests that the electorate as a whole may be much wiser than any individual voter.

The crucial problem with this mathematically elegant argument is that it does not work very well in practice.<sup>17</sup> Real voters' errors are quite unlikely to be statistically independent, as Condorcet's logic requires. When thousands or millions of voters misconstrue the same relevant fact or are swayed by the same vivid campaign ad, no amount of aggregation will produce the requisite miracle; individual voters' "errors" will not cancel out in the overall election outcome, especially when they are based on constricted flows of information (Page and Shapiro 1992, chaps. 5, 9). If an incumbent government censors or distorts information regarding foreign policy or national security, the resulting errors in citizens' judgments obviously will not be random.<sup>18</sup> Less obviously, even unintentional errors by politically neutral purveyors of information may significantly distort collective judgment, as when statistical agencies or the news media overstate or understate the strength of the economy in the run-up to an election (Hetherington 1996).

Bartels (1996) estimated how well the overall outcomes of six presidential elections matched what they would have been if every voter had been "fully informed." The average discrepancy between the actual popular vote and the hypothetical "fully informed" outcome of each election amounted to three percentage points—more than enough to swing a close contest. Related analyses of voting behavior in Sweden (Oscarsson 2007), Canada (Blais et al. 2009), Denmark (Hansen 2009), and many other countries (Arnold 2012) have found similar effects of information on aggregate election outcomes. Thus the lack of political knowledge matters—not only for individual voters, but also for entire electorates, the policies they favor, and the parties they elect.

#### THE ILLUSION OF "ISSUE VOTING"

The spatial theory of voting cast "issue proximity" as both the primary determinant of voters' choices and the primary focus of candidates' campaign

17 Formal theorists have also raised questions regarding the logical underpinnings of the argument, which typically hinge on the assumption that voters behave "sincerely" rather than strategically (Austin-Smith and Banks 1996; Feddersen and Pesendorfer 1998).

18 What we have in mind is that voters' errors can be correlated and thus not independently distributed—not the random white noise that Condorcet assumed. At the same time, voters' judgments typically depend on irrelevancies that do not reflect incumbent competence, so that election outcomes are not predictable by rational considerations—elections are "random" in that sense. Thus "nonrandom" (correlated) voter errors can lead to "random" (unpredictable from rational considerations) election outcomes.

strategies. Over the course of the 1960s and 1970s, this theoretical development was gradually but powerfully translated into empirical analyses of voting behavior. In the authoritative American National Election Studies conducted by the University of Michigan, questions regarding issues of public policy were increasingly recast as seven-point "issue scales" directly inspired by spatial theory. Survey respondents were invited to "place" themselves, candidates, and parties on each issue dimension. The proliferation of issue scales provided ample raw material for naive statistical analyses relating vote choices to "issue proximities" calculated by comparing respondents' own positions on these issue scales with the positions they attributed to the competing candidates or parties.

The causal ambiguity inherent in statistical analyses of this sort was clear to scholars of voting behavior by the early 1970s. Richard Brody and Benjamin Page (1972; Page and Brody 1972) outlined three distinct interpretations of the positive correlation between "issue proximities" and vote choices. The first, *policy-oriented evaluation*, corresponds to the conventional interpretation of issue voting in the folk theory of democracy: prospective voters observe the candidates' policy positions, compare them to their own policy preferences, and choose a candidate accordingly. The second, *persuasion*, involves prospective voters altering their own issue positions to bring them into conformity with the issue positions of the candidate or party they favor. The third, *projection*, involves prospective voters convincing themselves that the candidate or party they favor has issue positions similar to their own (and perhaps also that disfavored candidates or parties have dissimilar issue positions) whether or not this is in fact the case. In both the second and third cases, "issue proximity" is a *consequence* of the voter's preference for a specific candidate or party, not a *cause* of that preference.

Brody and Page (1972, 458) wrote, "We need some means for examining the potential for 'persuasion' and for 'projection' and of estimating them as separate processes. . . . If the estimation of policy voting is important to the understanding of the role of the citizen in a democracy—and theorists of democracy certainly write as if it is—then any procedure which fails to control for projection and persuasion will be an un dependable base upon which to build our understanding." Brody and Page's clear warning was followed by some resourceful attempts to resolve the causal ambiguity they identified (Jackson 1975; Markus and Converse 1979; Page and Jones 1979; Franklin and Jackson 1983). Unfortunately, those attempts mostly served to underline the extent to which the conclusions drawn from such analyses rested on fragile and apparently untestable statistical assumptions. Perhaps most

dramatically, back-to-back articles by Markus and Converse (1979) and by Page and Jones (1979) in the same issue of the *American Political Science Review* estimated complex simultaneous-equation models relating partisanship, issue proximity, and assessments of candidates' personalities using the same data from American National Election Studies surveys, but came to very different conclusions about the bases of voting behavior. If two teams of highly competent analysts asking essentially similar questions of the same data could come to such different conclusions, it seemed clear that the results of such exercises must depend at least as much on the analysts' theoretical preconceptions and associated statistical assumptions as on the behavior of voters.

In light of this apparent impasse, many scholars of voting behavior have preferred to sidestep the causal ambiguity plaguing the relationship between issue positions and votes by reverting to simple single-equation models in which issue positions can affect vote choices but not vice versa. In effect, they have relied on the assumptions of the folk theory of democracy rather than empirical evidence to resolve the problem raised by Brody and Page.<sup>19</sup> For example, Stephen Ansolabehere, Jonathan Rodden, and James Snyder (2008) cumulated responses to dozens of specific issue questions in American National Election Studies surveys into just two broad ("economic" and "moral") issue positions, discarding all of the remaining variation in specific issue responses as attributable to "measurement error." They imputed issue positions for voters who had none, or simply dropped them from the analysis. Then they imposed a model in which the observed relationship between issue positions and vote choices (net of partisanship and ideology) was attributed entirely to issue voting, with no allowance for persuasion or group identity effects. Unsurprisingly, they reported finding "stable policy preferences" and "strong evidence of issue voting" (Ansolabehere, Rodden, and Snyder 2008, 229).<sup>20</sup>

19 Some analysts have mitigated the resulting problem of bias due to *projection* by substituting sample average perceptions of the candidates' issue positions for individual respondents' own perceptions (e.g., Aldrich, Sullivan, and Borgida 1989; Erikson and Romero 1990; Alvarez and Nagler 1995). While this approach has the considerable virtue of reducing bias due to projection, it does nothing to mitigate bias due to *persuasion*; to the extent that voters adopt issue positions consistent with those of parties or candidates they support for other reasons, they will still (misleadingly) appear to be engaging in issue voting. Moreover, substituting sample average perceptions of the candidates' issue positions for respondents' own perceptions sacrifices a good deal of theoretical coherence, since it is very difficult to see how or why voters would compare their own issue positions to *other people's* perceptions of the candidates' positions, ignoring their own perceptions.

20 Freedler, Lenz, and Turney (2014) provided a reinterpretation of this evidence more consistent with Converse's (1964) view and our own.

Subsequent work by Gabriel Lenz (2009; 2012) provided substantial additional grounds for skepticism regarding inferences of this sort. Lenz used repeated interviews with the same individuals to show that *persuasion* plays a large role—and *policy-oriented evaluation* remarkably little role—in accounting for observed associations between issue positions and votes. That is, candidate choices determine issue positions, not vice versa.

In the 2000 presidential campaign, for example, candidate George W. Bush advocated allowing individual citizens to invest Social Security funds in the stock market, thereby catapulting a previously obscure policy proposal into the political limelight. Much of the news coverage and advertising in the final month of the campaign focused on the candidates' contrasting stands on the issue; in a typical "battleground" media market, the two candidates together ran about 200 ads touching on Social Security privatization just in the final *week* before Election Day (Johnston, Hagen, and Jamieson 2004, 153–159). And, sure enough, the statistical relationship between voters' views on Social Security privatization and their preferences for Bush or Al Gore (holding constant party identification) more than doubled over the course of the campaign.

This is exactly the sort of shift we might expect if voters were attending to the political debate, weighing the competing candidates' policy platforms, and formulating their vote intentions accordingly. However, Lenz's more detailed analysis employing repeated interviews with the same people demonstrated that this substantial increase in the apparent electoral impact of views about Social Security privatization was almost entirely illusory—due not to changes in vote intentions, but to Bush and Gore supporters *learning* their preferred candidate's position on the issue and then adopting it as their own. As Lenz (2012, 59) put it, "the increase in media and campaign attention to this issue did almost nothing to make people whose position was the same as Bush's more likely to vote for Bush than they already were."

On issue after issue—ranging from support for public works in 1976 and defense spending in 1980 to European integration in Britain to nuclear power in the Netherlands in the wake of the Chernobyl reactor meltdown—Lenz's analyses provided substantial evidence of *vote-driven* changes in issue positions but little or no evidence of *issue-driven* changes in candidate or party preferences. As John Zaller (2012, 617) put it, "Partisan voters take the positions they are expected as partisans to take, but do not seem to care about them." Lenz (2012, 235) characterized these findings as "disappointing" for "scholars who see democracy as fundamentally about voters expressing their views on policy," noting that the "inverted" relationship between issue

positions and votes seemed to leave politicians with "considerable freedom in the policies they choose."

## ELECTIONS AND PUBLIC POLICY

Almost all of the scholarly evidence we have considered thus far regarding public opinion and electoral behavior focuses on the attitudes and votes of individual citizens. Fortunately, we are not limited to individual-level analyses of public opinion and voting behavior. We can also attempt to assess directly how elections shape democratic politics. Does issue voting compel both parties to adopt policy positions close to those of the median voter, as the spatial theory of elections implies?

As we have seen, U.S. presidential elections in the post–World War II era—and especially the landslide defeats suffered by Barry Goldwater in 1964 and George McGovern in 1972—seemed to comport rather well with the predictions of the spatial theory. However, more systematic research on U.S. presidential elections has suggested that Goldwater and McGovern's losses had less to do with their issue positions than with election-year economic conditions; ideological "extremism" probably cost them just a few percentage points of the popular vote (Barrels and Zaller 2001; Cohen and Zaller 2012). More generally, the impact of candidates' policy stands on election outcomes—at least over the range of policy stands observed in modern presidential elections—seems to be quite modest. As Zaller (2012, 616) put it, "the penalty for extremism, if real, is not large."

The broad analysis of U.S. public policy in Robert Erikson, Michael MacKuen, and James Stimson's *The Macro Polity* (2002, 303–311) similarly underlines the failure of issue voting to discipline politicians in the manner suggested by the spatial theory of elections. Erikson, MacKuen, and Stimson measured the ideological tenor of policy activity in each branch of Congress and the White House over more than 40 years. They found that policy outcomes shifted substantially when partisan control shifted from Democrats to Republicans or from Republicans to Democrats. The public's "policy mood" (Erikson, MacKuen, and Stimson 2002, 194–205) also influenced policy regardless of which party was in control, but that effect was small by comparison. For example, the estimated impact on White House policy activity of moving from the most conservative "policy mood" recorded in four decades to the most liberal "policy mood" was only about one-third as large as the estimated impact of replacing a typical Republican president with a typical Democrat. The estimated effects of partisan control on congressional



policy activity were even larger. The implication is that citizens affect public policy—insofar as they affect it at all—almost entirely by voting out one partisan team and replacing it with another.

If the election of a Republican or Democratic president provided a reliable signal of the public's "policy mood," the resulting swing to right or left in policy outcomes might be characterized as a reflection of "majority rule" (though not in the sense suggested by the median voter theorem). The authors of *The Macro Policy* argued that presidential election outcomes are strongly affected by the public's "policy mood" (Erikson, MacKuen, and Stimson 2002, chap. 7). However, their statistical analyses required delicate controls for the prevailing balance of partisan loyalties in the electorate and the (inferred) ideological positions of the competing candidates. Subsequent analyses have found the apparent impact of "policy mood" evaporating once election-year economic conditions are taken into account (Cohen and Zaller 2012, table 3). Meanwhile, scholars attempting to forecast presidential election outcomes (e.g., Abramowitz 2012; Erikson and Wlezien 2012; Hibbs 2012) have generally been content to ignore "policy mood," issue preferences, and ideology—a telling indication that these factors are of relatively little importance in determining who wins.

Studies of Congress likewise find that the policy preferences of citizens in a given state or district are only modestly predictive of election outcomes—and that Democrats and Republicans routinely take very different stands once they are elected, even when they represent states or districts with very similar political views. Both of these points are clear in figure 2.1, which relates the overall roll call voting record of each member of the House of Representatives in the 112th Congress (2011–2013)<sup>21</sup> to the policy preferences of his or her constituents.<sup>22</sup> (Republican members are denoted by diamonds

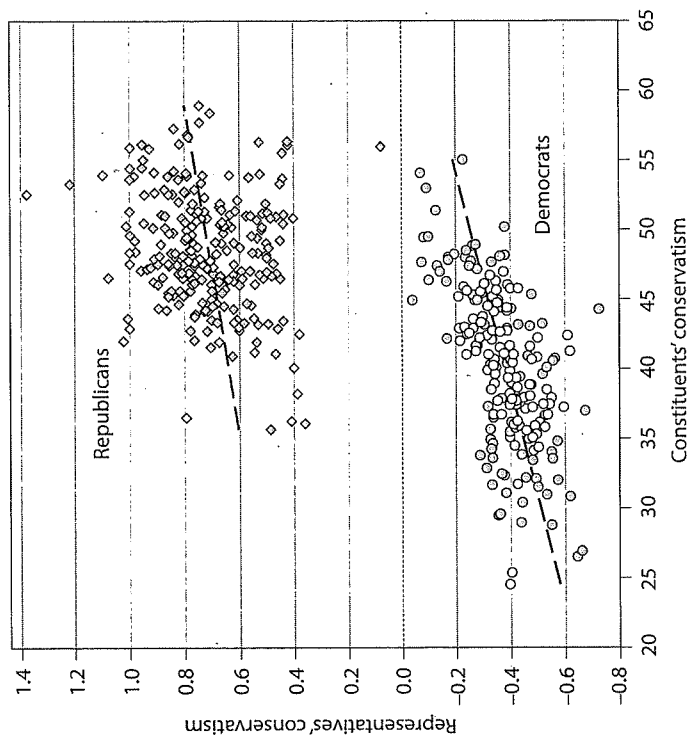


Figure 2.1. District Opinion and Representatives' Roll Call Votes, 2011–2013

and Democrats by circles.) District preferences are measured using a 12-item scale including liberal-conservative self-identification, beliefs about climate change, and support for the Affordable Care Act, domestic spending, the Iraq War, gays in the military, gun control, affirmative action, environmental protection, defense spending, a path to citizenship for illegal immigrants, and abortion.<sup>23</sup> The scale runs from 0 (most liberal) to 100 (most conservative), but the range of district averages is much narrower, from 24.5 to 59.0.

citizens (Vavreck and Rivers 2008). Additional information regarding the CCES surveys is available at <http://projects.iq.harvard.edu/cces/home>.

<sup>23</sup> Factor analysis produced a single dimension with factor loadings ranging from .78 for liberal-conservative self-identification to .50 for abortion. The resulting weights of the individual survey items are: .176 for liberal-conservative self-identification, .161 for beliefs about climate change, .104 for domestic spending, .083 for support for the Affordable Care

<sup>21</sup> We summarize each representative's entire roll call voting record using an index developed by Keith Poole and Howard Rosenthal (2007). Their (first-dimension) DW-NOMINATE scores represent "ideal points" that account as accurately as possible for each representative's entire roll call voting record under the assumption of spatial voting. The scale on which the ideal points are measured is arbitrary, but they are conventionally normalized to run from -1 for the most liberal member of the House to +1 for the most conservative member. However, due to constraints imposed by the DW-NOMINATE algorithm on the movement of each representative on the scale from one Congress to the next, the range of actual scores in figure 2.1 is from -0.729 to 1.376.

<sup>22</sup> These data are from surveys conducted in 2010 and 2012 by the Internet survey firm YouGov as part of the Cooperative Congressional Election Study (CCES). There were a total of 52,464 respondents in 2010 and 51,661 in 2012. The combined sample size in each congressional district ranged from 88 to 515 and averaged 239. YouGov employs opt-in recruiting, but uses matching and weighting to produce representative samples of adult U.S.



The most liberal congressional districts in the country, at the far left of figure 2.1, invariably elected Democrats to the House in 2010. (These were overwhelmingly urban and mostly majority-minority districts.) At the opposite extreme, the most conservative districts almost all elected Republicans. However, for districts in the broad middle of the political spectrum, election outcomes were a rather unreliable reflection of citizens' policy preferences. Moderately liberal districts (in the second quartile of the national distribution) elected Republicans 46% of the time, while moderately conservative districts (in the third quartile) elected Democrats 25% of the time.<sup>24</sup>

The modest correlation between constituents' preferences and election outcomes implies substantial variation in representation, given the gulf in roll call voting behavior between Republicans and Democrats representing similar districts in figure 2.1. Nor is it the case that representatives won election in what looked like uncongenial districts by catering closely to citizens' preferences at the expense of their own (or their parties') convictions. The dotted lines in figure 2.1 summarize the separate linear relationships between the conservatism of each party's representatives and their constituents' preferences. Within each party, there is a modest positive relationship between constituents' preferences and House members' roll call votes. However, the magnitudes of those relationships are dwarfed by the distance between the two lines, which represents the expected difference in conservatism between Republican and Democratic members representing districts with identical public opinion.

Clearly, Republican and Democratic members of Congress representing constituents with similar preferences behaved in very different ways. Whether these differences were produced by differences in the representatives' personal ideological convictions or party pressures or other factors is, for our purposes here, irrelevant. The key point is that representatives' voting behavior was not strongly constrained by their constituents' views.<sup>25</sup> Elections do *not* force

Act, .076 for environmental protection, .072 for affirmative action, .067 for gun control, .059 for the Iraq War, .058 for defense spending, .054 for gays in the military, .051 for abortion, and .039 for a path to citizenship for illegal immigrants.

<sup>24</sup> The corresponding relationship between U.S. Senate election outcomes and constituents' policy preferences is even weaker. The difference may reflect the fact that Senate elections tend to involve more publicity and campaign spending than House elections, making voters more susceptible to being swayed by candidate-specific factors unrelated to policy (Krasno 1997).

<sup>25</sup> Since the analysis presented in figure 2.1 employs incommensurate measures of representatives' voting behavior and constituents' preferences, we are not able to say whether

successful candidates to reflect the policy preferences of the median voter, as Downsian logic implies.

The pattern of partisan polarization evident in figure 2.1 is not a fluke attributable to a particular congressional session or opinion survey.<sup>26</sup> Indeed, a historical analysis of every Congress going back to the 1870s (using presidential election returns as proxies for constituents' preferences) suggests that similar differences in expected roll call voting patterns between Republicans and Democrats representing similar constituencies have been fairly common in the past 140 years (Bartels, Clinton, and Geer forthcoming).<sup>27</sup>

Scholars of comparative politics sometimes argue that the pattern of alternating partisan extremism that has characterized the American political system through much of its history is absent or attenuated in multiparty systems where legislative seats are allocated through proportional representation. According to G. Bingham Powell (2000, 243), for example, "proportional systems are more successful in getting governments (and even more so the influential policymakers) close to the median citizen." However, assessments of this sort depend on untested assumptions about how parties' positions get translated into policies under different institutional arrangements. More direct assessments of patterns of responsiveness have found little consistent difference between proportional and majoritarian systems in their extent of correspondence between median voters' preferences and actual policy outcomes (e.g., Kang and Powell 2010; Bartels 2015). Perhaps some other institutional arrangement would work better; we do not know. But at present, there is little evidence to suggest that changes in electoral institutions will be sufficient to ensure popular control of public policy through electoral competition.

## CONCLUSION

The folk theory of electoral democracy—the notion that elections can “reveal the will” or the preferences of a majority on a set of issues,” as Dahl (1956,

Republican representatives were more conservative than their constituents or Democratic representatives were more liberal or both.

<sup>26</sup> For example, Bartels (2008, 256) documented a qualitatively similar pattern for the U.S. Senate in the late 1980s and early 1990s, and Joshua Clinton (2006, 401) did the same for the U.S. House of Representatives in 1999–2000.

<sup>27</sup> Bartels, Clinton, and Geer's (forthcoming) analysis suggests that the intense partisan polarization of congressional roll call voting since 1994 (above and beyond what could be accounted for by differences in district preferences) was matched in the period from 1874 through 1920. By comparison, the period from the mid-1930s through the mid-1970s was one of consistently low partisan polarization by this measure.

131) put it—has played a central role in both political science and popular thinking about politics. Scholars have elegantly codified the populist ideal in the “spatial theory” of elections. The behavior it enshrines as normative, “issue voting,” is widely regarded as a hallmark of good citizenship. Indeed, a conscientious voter nowadays can choose among a variety of websites inviting her to answer a series of policy questions and be told which candidates to support—apparently on the assumption that nothing other than the candidates’ policy positions (and just these policy positions) should matter to her.<sup>28</sup> The social science theorizing and the cultural norm are both derived directly from the folk theory of democracy.

Unfortunately, as we have seen, this populist ideal in both its scientific and popular incarnations suffers from grave logical and practical problems. Both the remarkable theoretical insights of Arrow (1951) and his successors and the seminal empirical research of Converse (1964) and many others punched significant holes in the romantic populist notion of democracy. Although a great deal of subsequent scholarly effort has been devoted to recasting, circumscribing, or rejecting their claims, repeated attempts to sidestep the theoretical and empirical deficiencies of the populist ideal have failed, leaving the “strong challenge to democratic hopes” (Kinder 2003, 15) posed by modern social science fundamentally intact.

These scientific findings have had little effect on practical politics. Joseph Schumpeter (1942, 250) argued—perhaps wishfully—that “today it is difficult to find any student of social processes who has a good word for” the simplistic notions of the folk theory. Nevertheless, he added, “action continued to be taken on that theory all the time it was being blown to pieces. The more untenable it was being proved to be, the more completely it dominated official phraseology and the rhetoric of the politician” (Schumpeter 1942, 249). More than seven additional decades of demolition work have done little to alter that picture.

Indeed, periodic frustration with the apparent failure of elections to faithfully translate “the will of the people” into public policy has prompted repeated attempts to wrench American political practice into closer accordance with the folk theory. The principal goal of these efforts has been to constrain or even bypass those whom the reformers blamed for their disappointments—professional politicians. Perhaps most radically, reformers in the Progressive

28 See, for example, <http://votesmart.org/voteeasy/>, [http://www.votehelp.org/](http://www.ontheissues.org/http://www.votehelp.org/), <http://selectsmart.com/politics.html>, <http://www.quizrocket.com/political-party-quiz>.

Era attempted “to restore the absolute sovereignty of the people” (Bourne 1912) by circumventing the traditional electoral process altogether, instituting “direct democracy” via initiatives and referendums. We turn next to the promise and pitfalls of these repeated attempts to impose an idealistic theory on the recalcitrant reality of American democracy.

considerable enthusiasm—if not by political philosophers or ordinary citizens, at least by political scientists who spent their time studying public opinion, elections, and the policy-making process. The underlying idea was not new, but its explicit and detailed formulation as a defense of democracy brought it fresh respectability.

The key insight of this alternative theory of democracy was that voters could exert substantial control over their leaders, despite knowing little about the details of public policy, simply by assessing the *performance* of incumbent officials, rewarding success and punishing failure. In one of the first and most influential formulations of this perspective, V. O. Key (1966, 61) portrayed “the electorate in its great, and perhaps principal, role as an appraiser of past events, past performance, and past actions. It judges retrospectively; it commands prospectively only insofar as it expresses either approval or disapproval of that which has happened before.”

This *retrospective theory* of political accountability seems to provide a compelling way to think about the relationship between leaders and citizens in democratic political systems. Empirically, it accounts for fluctuations in the electoral fortunes of incumbent leaders and parties much more successfully than any spatial theory of issue voting. Moreover, it rescues political leaders from their fate as hapless automatons “converging” on the policy preferences of the median voter in the spatial theory—or, perhaps more realistically, as demagogues pretending to cater to the garbled voice of the people. In Key’s words (1958, 590), “the most acute ear attuned to the voice of the people can sense only the vaguest guidance for innovation to cope with the questions that must be met day by day as an Administration governs. The efficacy of self-government thus depends on party and governmental leadership with the initiative and imagination necessary to meet the public problems that develop and with the courage to assume the political risks involved.” The retrospective theory gives political leadership its due, leaving initiative in the hands of elected officials to further voters’ well-being by any feasible and legitimate means at their disposal.

The normative appeal of the retrospective theory stems in no small part from the fact that it seems to save voters from the charge that they are too uninformed or too disengaged to play a meaningful role in the democratic process. For example, Morris Fiorina (1981, 5) argued that retrospective voters “need *not* know the precise economic or foreign policies of the incumbent administration in order to see or feel the *results* of those policies. . . . In order to ascertain whether the incumbents have performed poorly or well, citizens need only calculate the changes in their own welfare. If jobs have been lost in a recession, something is wrong. If sons have died in foreign rice paddies,

#### CHAPTER FOUR

## A Rational God of Vengeance and of Reward? The Logic of Retrospective Accountability

What is more dramatic than the spectacle we have seen repeated, and doubtless long shall see—the popular judgment taking the successful candidates on trial in the offices—standing off, as it were, and observing them and their doings for a while, and always giving, finally, the fit, exactly due reward?

—Walt Whitman, *Democratic Vistas* (1871, 31)

To support the Ins when things are going well; to support the Outs when they seem to be going badly, this, in spite of all that has been said about tweedledum and tweedledee, is the essence of popular government.

—Walter Lippmann, *The Phantom Public* (1925, 126)

The developments in survey research, political psychology, and social choice theory summarized in chapter 2 severely undercut the intellectual foundations of the folk theory of democracy. Fourth of July rhetoric notwithstanding, the hope that elections could ensure “the continued responsiveness of the government to the preferences of its citizens” (Dahl 1971, 1) seemed to many students of democracy increasingly hollow in light of mounting evidence regarding voters’ opinions and behavior. Absent some reliable mechanism for eliciting meaningful popular preferences and aggregating them into a coherent policy prescription, what, concretely, should a “responsive” democratic government actually do?

No wonder, then, that when a less demanding version of democratic theory began to draw renewed attention in the 1960s, it was greeted with

something is wrong. If thugs make neighborhoods unsafe, something is wrong. If polluters foul food, water, or air, something is wrong." By dispensing with the unrealistic notion that ordinary citizens vote on the basis of detailed preferences regarding every issue that might conceivably come before their future leaders, the theory of retrospective voting made the rather bleak portrait of poorly informed, habitual political behavior provided by the survey researchers of the 1950s and 1960s, if not quite irrelevant, at least not fatal. Voters need only monitor their own and their fellow citizens' day-to-day experiences and well-being.

The seminal works of Key (1966), Fiorina (1981), and Gerald Kramer (1971) all portrayed retrospective voting as evidence of the fundamental rationality of American voters and elections. For example, Kramer (1971, 140) characterized his analysis as demonstrating that election outcomes "are not 'irrational,' or random, or solely the product of past loyalties and habits, or of campaign rhetoric and merchandising." Fiorina (1981, 4) insisted that American democracy "functions better than one would expect, given behavioral political science's destruction of old illusions," and discerned "rough justice" in the electoral punishment meted out to incumbent politicians by American voters in the elections of 1932, 1952, 1958, 1968, and 1974. And Key (1966, 7) famously put the "perverse and unorthodox argument" of his book in a nutshell: "voters are not fools."

As often happens, a coherent account of retrospective voting within the framework of rational choice theory (Ferejohn 1986; Rogoff 1990; Banks and Sundaram 1993; Fearon 1999) came along only after considerable empirical evidence establishing the substantive importance of the phenomenon had already piled up. Nevertheless, the development of coherent accounts of why "rational" citizens might be expected to engage in retrospective voting—and might benefit from doing so—added considerably to the theory's normative appeal. In fact, two distinct accounts of that kind were produced. Much of the present chapter is devoted to summarizing them.

While we attach great importance to the realism, empirical power, and normative appeal of the retrospective theory, we believe that its implications for democracy are less unambiguously positive than existing literature tends to suggest. In particular, the theory as it stands fails to do justice to the very considerable logical and informational difficulties faced by retrospective voters attempting to assess "changes in their own welfare," as Fiorina put it, and to translate those assessments into meaningful evaluations of incumbent political leaders.

For one thing, it is by no means obvious that voters can "ascertain whether the incumbents have performed poorly or well" simply by calculating "changes

in their own welfare," as Fiorina claimed. If jobs have been lost in a recession, something is wrong, but is that the president's fault? If it is not, then voting on the basis of economic conditions may be no more sensible than kicking the dog after a hard day at work.<sup>1</sup> An even more fundamental problem is that voters may have great difficulty accurately assessing "changes in their own welfare." Proponents of retrospective voting have routinely assumed that voters *know* when "thugs make neighborhoods unsafe" or "polluters foul food, water, or air," as Fiorina put it. But that is by no means obvious. To the extent that voters' assessments of their own well-being are erroneous, retrospective voting will succeed much less well in selecting good leaders and in disciplining them to pursue voters' interests.

We believe that these difficulties are more serious than previous scholars of retrospection have generally thought. As a result, we are left with a much more measured view of the extent to which retrospective voting justifies "a portrait of citizens moving to considered decision as they play their solemn role of making and unmaking governments" (Key 1966, 4).

#### EVIDENCE OF RETROSPECTIVE VOTING

Politicians and political observers have long believed that voters punish failure and reward success, especially in the management of the economy.<sup>2</sup> President James Buchanan blamed his party's dismal showing in the 1858 midterm election on economic distress resulting from the Panic of 1857 (Huston 1987, 166–168). And no less a politician than Benjamin Disraeli recognized the electoral significance of bad harvests in 19th-century Britain.<sup>3</sup> In 1879, he wrote to a colleague that "the only danger and difficulty which the present Ministry has to encounter are natural. . . . After four bad harvests in this country, we are apparently about to meet a fifth dearth" (Monypenny and Buckle 1929, 1347). Sure enough, in a disastrous election the following spring the Tory government fell.

The spatial model of party competition outlined in chapter 2 provides no help in understanding politics of this sort. As Donald Stokes (1963, 373)

1 Fiorina (1981, 202, emphasis original) noted in passing that the normative appeal of the theory of retrospective voting hinges in significant part on the assumption "that the *electorate does a passable job of attributing responsibility to government decision makers*." However, his brief discussion of the problems raised by that assumption focused almost entirely on the issues of divided government and responsible parties, rather than on the more basic questions considered here.

2 William Keech (2013) provided a comprehensive review of theory and evidence related to "pocketbook voting."

3 We are indebted to W. Phillips Shively for this reference.

argued in his insightful critique of Anthony Downs's version of the spatial model, "throwing the rascals out is very different from choosing between two or more parties on the basis of their advocacy of alternatives of government action." Stokes insisted upon the electoral significance of "valence-issues . . . that merely involve the linking of the parties with some condition that is positively or negatively valued by the electorate" in contrast to "position-issues" of the sort represented in Downs's model: "in American presidential elections of the past generation it is remarkable how many valence-issues have held the center of the stage," from depression and recovery in the 1930s and 1940s to the Korean War and corruption in 1952 to America's international prestige in 1960.

Political scientists in the subsequent half century have provided a great deal of more systematic evidence regarding the political significance of Stokes's "valence-issues." Key's evidence in *The Responsible Electorate* consisted primarily of statistical analyses of the correlates of vote-switching from one presidential election to the next. For example, in each of Franklin Roosevelt's reelections, poorer people and blue-collar workers—the presumed beneficiaries of New Deal policies—were less likely than wealthy people and those in business and professional occupations to report having switched their votes from Democrats to Republicans (Key 1966, 35, 37). Moreover, people who reported supporting specific policies of the incumbent administration—compulsory old-age insurance in 1936, the Wagner Labor Act in 1940, the Korean War in 1952—were less likely to defect than those who reported opposing those policies (Key 1966, 43, 47, 75, 97). And, most important for Key's argument, people who approved of the incumbent president's performance were more likely to "stand pat," even when the incumbent himself was not on the ballot, as in 1960 (Key 1966, 139).<sup>4</sup>

In *Retrospective Voting in American National Elections* (1981), Fiorina elaborated and extended Key's analyses in a variety of ways. He incorporated a wider variety of retrospective assessments—ranging from presidential

performance to avoiding war to personal financial circumstances—in statistical analyses of vote choices in a dozen presidential and congressional elections (Fiorina 1981, 35–43). He used repeated interviews with the same survey respondents to document the impact of retrospective evaluations on shifts in partisan identification (Fiorina 1981, 94–102). And he examined the extent to which "mediated retrospective evaluations"—assessments of incumbent performance—were grounded in "simple retrospective evaluations"—assessments of economic and other conditions as good or bad (or improving or deteriorating) without reference to government, political parties, or the president (Fiorina 1981, 108–129).

Scholars following in the footsteps of Key and Fiorina have generated a substantial body of research relating individuals' expressed retrospective judgments to their vote choices. Most of this work focuses specifically on judgments of national economic conditions or personal economic circumstances or both (Kinder and Kiewiet 1981), rather than on the broader variety of retrospective assessments examined by Key and Fiorina. Although most of the evidence comes from the United States, similar patterns of retrospective voting have been documented in many other countries (Lewis-Beck 1988; Duch and Stevenson 2008).

Virtually all of this work is subject to much the same methodological criticism we raised with respect to empirical studies of "issue voting" in chapter 2: observed cross-sectional relationships between retrospective evaluations (for example, assessments of incumbent performance or perceptions of the economy) and vote choices may reflect the electoral impact of retrospective evaluations, but they may also reflect the extent to which retrospective evaluations are constructed to *rationalize* vote intentions whose real causes lie elsewhere.<sup>5</sup> Indeed, as Kramer (1983) pointed out, in a single snapshot survey *real* national economic conditions are a constant, so *all* of the observed variation in *perceived* national economic conditions must reflect partisan bias,

4 While Key emphasized the electoral significance of retrospective judgments of incumbent policies and performance, he clearly supposed that voters were forward-looking as well as backward-looking. Indeed, he reported even stronger statistical relationships between *prospective* judgments and vote choices than between *retrospective* judgments and vote choice. In Gallup data from 1960, for example, vote-switchers overwhelmingly rated the party they supported as "best for people like yourself," more likely to keep the United States out of World War III, "better at keeping the country prosperous," and better at handling the "most important" specific problem facing the country (1966, 124, 125, 127, 133). Obviously, the causal status of judgments of this sort is far from clear. Did voters in 1960 support Nixon because they thought he would keep the country out of war, or did they think he would keep the country out of war because they already supported him?

5 Key clearly recognized this possibility, but downplayed its practical importance on grounds that seem distinctly tangential. The "relationship between policy outlook and vote," he wrote (1966, 45–46), "doubtless reflected to a degree the tendency of a voter on a specific question to improvise policy views that seem to be consistent with the way he planned to vote for other reasons entirely. A steadfast Democratic partisan might have been expected to opine that the Roosevelt administration has done a good job in handling the farm problem, if the question were put to him in that form. Yet, however such opinions come into being, their supportive function in the political system should be the same." While it is true that even "rationalizing" retrospective judgments may have a "supportive function," the fact that they are not genuine *causes* of voting behavior seems to us to be the key point.



overgeneralization from personal experience or local conditions, and other vagaries of individual perception.

As in the case of issue voting, it is possible to get some traction on this possibility by analyzing data from repeated surveys of the same individuals. Gabriel Lenz's analyses along these lines provided a striking contrast between the apparent causal impact of voters' policy preferences on one hand and their performance evaluations on the other. As we noted in chapter 2, Lenz (2012, 216) found "surprisingly little evidence" that policy preferences "carried much weight in voters' judgments" once he allowed for the reciprocal impact of vote intentions on policy preferences. However, his parallel analyses of performance evaluations provided "further evidence of the importance of performance-related issues, such as the economy. Although politicians, journalists, and scholars often interpret elections ideologically," Lenz concluded (2012, 225), "a growing body of evidence suggests that it is the economy and other performance domains, not ideology, that largely explain election outcomes. My results support this view."

In the case of economic voting, the ready availability of objective economic data provides additional leverage for disentangling the causal relationship between retrospective assessments and vote choices. While economic perceptions and evaluations may be colored by politics, income and unemployment statistics have an independent reality that may (or may not) be manifested in voting behavior. Insofar as it is, the connection cannot simply be a matter of voters rationalizing choices unrelated to the actual condition of the economy. Thus, systematic analyses of the aggregate-level relationship between economic conditions and election outcomes may complement and bolster analyses showing that individual voters' economic perceptions shape their electoral behavior.<sup>6</sup>

Harold Gosnell and various coauthors produced a series of remarkably sophisticated and politically acute studies of the relationship between economic conditions and election outcomes in the New Deal era (Gosnell and Gill 1935; Gosnell and Schmidt 1936; Gosnell and Pearson 1939; Gosnell

6 Of course, the aggregate-level relationship between objective economic conditions and election outcomes need not be attributed solely to retrospective voting. For example, economic growth may inspire wealthy campaign donors to contribute to the incumbent's reelection campaign, providing resources to buy votes without the voters themselves knowing or caring about the state of the economy (Bartels 2008, 116–122). However, the strength of the relationships between economic conditions and aggregate economic perceptions and between economic perceptions and voting behavior leaves little doubt that most of the aggregate-level relationship is indeed attributable to retrospective voting of one sort or another.

and Cohen 1940; Gosnell and Coleman 1940; Gosnell 1942). Using an impressive variety of economic and political data from Chicago, Iowa, Pennsylvania, and Wisconsin, and statistical tools that would not come into general use in the social sciences until a quarter century later, they demonstrated that voters' recent economic circumstances did indeed appear to influence election outcomes.

The line of research advanced by Gosnell made little further progress over the next few decades, as scholars of electoral politics turned their attention to the exciting new possibilities for detailed individual-level analysis offered by the advent of scientific opinion surveys. However, the study of economics and elections was reinvigorated by the publication of Kramer's 1971 article on "Short-Term Fluctuations in U.S. Voting Behavior, 1896–1964." Whereas Gosnell and his colleagues had focused on single states or localities over relatively short periods of time, Kramer analyzed the relationship between national economic conditions and the national congressional vote across 34 biennial elections.<sup>7</sup> He examined the electoral relevance of a variety of economic indicators, including real income, unemployment, and inflation, as well as the effects of congressional incumbency and presidential coattails. He concluded that economic fluctuations—most notably, election-year changes in real per capita income—are important influences on congressional elections, . . . account[ing] for something like half the variance of the congressional vote, . . . over the period considered" (Kramer 1971, 140–141).

Scores of scholars following in Kramer's wake have examined the impact of economic conditions on election outcomes in the United States and elsewhere.<sup>8</sup> While various aspects of the relationship—including the magnitude and timing of electoral responses to specific economic conditions, the psychology linking economic conditions to voters' choices, and the implications of institutional and contextual variation—are far from settled, a virtual

7 Kramer (1971, 135) regarded congressional elections as providing the clearest possible test of retrospective theory, "since of the races for national office, House contests come closest to the Downsian case of relatively anonymous candidates competing as members of a common party team." His statistical analysis suggested that presidential elections "are substantially less responsive to economic conditions" (Kramer 1971, 141). However, subsequent analyses—most of which have focused on the post-World War II era—have consistently found larger effects of economic conditions in presidential elections than in congressional elections (for example, Erikson 1989, 1990).

8 Lewis-Beck and Stegmaier (2007) provided a broad survey of relevant literature. For US. presidential elections, Bartels and Zaller (2001) compared a variety of alternative measures of economic performance and probed the robustness of the statistical results to variations in model specification.



consensus has emerged that the electoral impact of economic conditions is real and substantial.

#### RETROSPECTIVE VOTING AS A MECHANISM FOR SELECTING COMPETENT LEADERS

The most straightforward way to interpret retrospective voting is as an attempt by voters to select the best available team of political leaders. In this view, there are consequential differences among the competing parties in any given election—differences in motivation, competence, ideology, or some combination of these and other factors. These differences imply differences in future personal or collective well-being, at least in expectation.<sup>9</sup> The voters' problem is to forecast future well-being under each of the competing parties and choose the one offering the most favorable prospects.

This interpretation of retrospective voting portrays voters as rational and forward-looking; but their *prospective* choices are rendered *retrospective* through the auxiliary assumption that the parties' past performance in office can generate rational expectations about future performance. In the influential formulation of Anthony Downs (1957, 106, 39), "rational behavior is impossible without at least some way of forecasting future events. . . . Since one of the competing parties is already in power, its performance in period  $t$  gives the best possible idea of what it will do in the future, assuming its policies have some continuity."

While the basic logic of retrospective selection is straightforward, the precise implications for voting behavior depend on the nature of the presumed relationship between past performance and future prospects. Assuming for the sake of argument that the voter's well-being under the incumbent party in the most recent period provides "the best possible idea" of its future performance, how informative is that? And how informative is earlier experience with the incumbents, or with the "out" party? Does the diagnostic value of past experience vary with turnover in personnel or changes in social conditions?

The appendix to this book sets out a simple mathematical model of retrospective selection providing a framework for addressing questions of this

9 Political scientists have generated a substantial literature focusing on the distinction between "sociotropic" assessments of the national economy and "pocketbook" assessments of personal economic well-being (Kinder and Kiewit 1981; Kramer 1983; Markus 1988). For our purposes, it does not matter whether voters define well-being individually or collectively; any consistent desideratum will suffice to ground a retrospective theory of democracy.

sort. The model assumes that a representative voter's *electorally relevant subjective well-being* under the incumbent is determined by a simple combination of the incumbent's intrinsic competence and random factors outside the incumbent's control—good or bad luck.<sup>10</sup> The problem, from the voter's perspective, is to infer as much as possible about the incumbent's intrinsic competence, and thus about her likely performance in the next period if she is reelected.

If the incumbent's intrinsic competence does not vary over time, or varies purely randomly with no correlation across time periods, the best available estimate of her competence is simply the average well-being experienced under her leadership in the current and previous periods. An important implication of this fact is that a rational voter's assessment of the incumbent's competence will take into account performance in every relevant period, not just in the current period. In Fiorina's (1981, 84) felicitous phrase, the rational voter maintains a "running tally" of subjective well-being under the incumbent, updating it continually on the basis of good or bad experience. Thus, retrospection provides a basis not only for vote choices in a given election but also for more or less durable partisan preferences—what Fiorina (1977) called "a political theory of party identification."

If we relax the assumption that the incumbent's competence is constant over time, then our voter should form his forecast of the incumbent's future competence somewhat differently, depending upon exactly how competence is supposed to evolve over time (Gerber and Green 1998; Achen 2012). In general, however, the possibility of changes in competence makes previous experience less relevant, leading voters to attach relatively more weight to more recent experience.

The efficacy of the voter's effort to select a competent incumbent in this framework obviously depends on the quantity and quality of evidence provided by prior experience. If the continuity of party leadership is minimal, a party's past performance may have little bearing on how well it will perform in the future. If voters' subjective well-being is determined more by extraneous

10 We focus here on a single representative voter. More generally, one might posit a parallel relationship for each voter, with individual-specific well-being determined by individual-specific incumbent competence and individual-specific doses of luck. Doing so would add empirical plausibility at the cost of considerable strategic complexity. As Hibbs (2006, 570) noted, if voters' retrospective judgments are heterogeneous "incumbents could pursue a divide and rule strategy by exploiting distributive conflicts in the electorate, and thereby mitigate, or perhaps avoid completely, the discipline of having to satisfy a minimal standard of macroeconomic performance augmenting aggregate welfare."

events such as droughts or oil price shocks than by the quality of the incumbent's policies and management, electoral rewards and punishments may often be misdirected. However, to the extent that voters' past political experience does shed light on the likely quality of future leadership, they can and should weigh that experience when they go to the polls.

#### RETROSPECTIVE VOTING AS A MECHANISM FOR SANCTIONING LEADERS

The logic of retrospective voting as a mechanism for selecting competent leaders is fairly straightforward. However, it may be more surprising that retrospective voting can promote voters' well-being even when there are *no differences in competence* among potential leaders. As long as incumbents (or their parties) desire reelection, the knowledge that voters at the next election will reward them for success or punish them for failure provides incentives for leaders to maximize voters' well-being as best they can. As Fiorina (1981, 11) put it, "Given political actors who fervently desire to retain their positions and who carefully anticipate public reaction to their records as a means to that end, a retrospective voting electorate will enforce electoral accountability, albeit in an *ex post*, not an *ex ante*, sense."

To explicate this logic, the appendix sets out a second model of retrospective voting in which variation in incumbent competence plays no role (Ferejohn 1986). Incumbents and challengers are assumed to be identical; the issue of political accountability is a simple matter of "moral hazard" of the sort familiar from economic models of principal-agent relationships (Laffont and Martimort 2002). The current incumbent may be tempted to focus on other goals besides maximizing voters' subjective welfare,<sup>11</sup> and voters cannot directly monitor how diligently the incumbent is working on their behalf. However, they can tell when they experience pain or pleasure, and they know that their pain or pleasure is attributable, in part, to the incumbent's efforts.

In contrast to the retrospective selection model, the *sanctioning* model assumes that each voter's electorally relevant subjective well-being depends solely on the amount of effort expended by the incumbent leader on the

11 In the theoretical literature, this is often labeled "shirking." We think the image conjured up by that label, of politicians sitting around with their feet up on their desks, is inappropriate, models of this sort can be used to explore the implications of a much broader class of competing goals that might tempt incumbent leaders to deviate from doing whatever will maximize their prospects for reelection—including graft, courageous leadership on behalf of their constituents' long-term welfare, or anything in between.

voter's behalf and on the net effect of random forces beyond the incumbent's control. "Luck" plays much the same role in this model as in our model of selection on the basis of incumbent competence, representing the impact on the voter's well-being of factors unrelated to the incumbent's effort. However, we assume that the voter cannot directly observe the incumbent's expenditure of effort or distinguish the effects of effort and luck; thus, his decision whether or not to return the incumbent leader to office for the next period is governed by the sum of both factors.

In this second model, the incumbent knows that her probability of reelection depends in part on how much effort she exerts on the voters' behalf. However, she also knows that her fate depends in part on good or bad luck. Moreover, since effort devoted to increasing the voter's subjective well-being is presumed to be costly—it may require foregoing graft, ideological goals, or the esteem of future historians—incumbents weigh the opportunity cost of exerting effort on the voters' behalf against the potential benefit of reelection.

For their part, voters cannot directly control or even observe the incumbent's exertion of effort. However, they can decide whether or not to reelect the incumbent based on good or bad outcomes attributable, in part, to the incumbent's level of effort. They want to choose a standard for reelection that will elicit as much effort as possible from incumbents. Intuitively, one might think that they should set the standard as high as possible, since their expected well-being is always higher when incumbents exert more effort on their behalf. However, setting the standard too high actually discourages effort; if incumbents know they are unlikely to get reelected whatever they do, they are more likely to shirk. Indeed, in equilibrium, it turns out that voters should always set the standard for reelection so that the incumbent's *ex ante* probability of reelection will be .5, regardless of whether that standard turns out to induce incumbents to exert a little or a lot of effort on their behalf. Thus, an interesting implication of the model is that we should expect incumbents to be reelected half the time.<sup>12</sup>

Nevertheless, the voters' expected well-being is greater when their assessment of the incumbent's performance depends less on luck and more on incumbent effort, since in that case the incumbent can expect a closer correspondence between exertion and reward (in the form of reelection). In that

12 David Mayhew (2008) has noted that, historically, incumbent U.S. presidents running for reelection have succeeded approximately two-thirds of the time. That success rate can be derived formally from a model that allows for differences in perceived competence between incumbents and challengers (Achen 2016), unlike the model we discuss here.

respect, the primary implication of the model focusing on variation in incumbent effort parallels that of the model focusing on variation in incumbent competence: Retrospective voting can be a powerful mechanism for electoral accountability, but only insofar as voters can discern and set aside irrelevant factors contributing to their subjective well-being.

#### CHALLENGES TO THE EFFECTIVENESS OF RETROSPECTIVE ACCOUNTABILITY

Jonathan Bendor, Sunil Kumar, and David Siegel (2010, 31) have argued that "it is important to start with a model of retrospective voting that is consistent with the strongest patterns uncovered by empirical students of voters: what voters know and how they think." We heartily agree, and the models of retrospective voting set forth in the appendix do just that. They focus (as voters mostly do) on ends rather than means, and they allow for the fact that voters' assessments of individual and collective well-being will only imperfectly reflect the competence and effort of political leaders.

The positive implication of these models is that retrospective voting can promote democratic accountability in two distinct ways—by selecting good leaders and by inducing leaders to exert effort on voters' behalf.<sup>13</sup> In that sense, our analysis validates the informal logic of Key, Fiorina, and other proponents of retrospective voting. However, our analysis also underlines the fact that the likely effectiveness of retrospective voting in promoting democratic accountability hinges crucially on the magnitude of random forces influencing voters' electorally relevant subjective well-being. While this is a common feature of formal models of retrospective voting, previous analysts have given it little attention. We explore it briefly here, and more expansively in chapters 5 to 7.

In the simplest version of our model of retrospective selection, which assumes that the competence of the incumbent is constant over time, the un-

13 We have presented separate models explicating the logic of these two distinct mechanisms of accountability. Extending our analysis to deal simultaneously with differences in intrinsic competence and incentive effects within a unified model would make the analysis more complicated and the results sensitive to arbitrary choices of assumptions regarding how the various moving parts fit together. James Fearon (1999) provided an insightful examination of the interaction of "selecting" and "sanctioning" processes. In his framework, "any variation in politicians' attributes or propensities relevant to their performance in office" gives voters incentives "to focus *completely* on choosing the best type when it comes time to vote" (Fearon 1999, 77); thus, the dual model reduces in practice to a model of selection. Jane Mansbridge (2009) provided a broader discussion of the history and implications of the "selection model" of political representation.

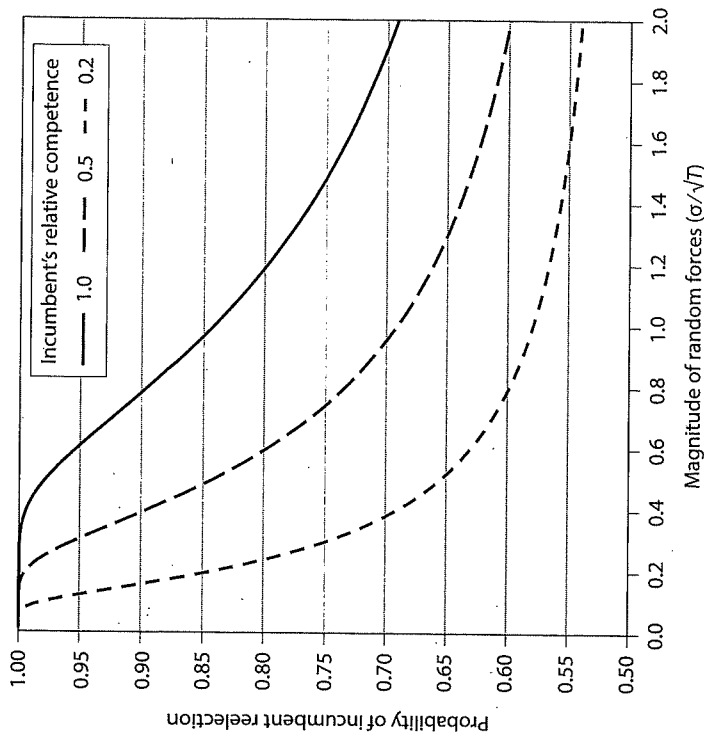


Figure 4.1. Blind Retrospection Impairs Voters' Judgment of Good Incumbents

certainty of the voter's assessment of that competence depends upon the magnitude of discrepancies between competence and subjective well-being and (inversely) upon the extent of the voter's relevant experience.<sup>14</sup> The implications of that uncertainty are illustrated in figure 4.1, which shows the probability of incumbent reelection as a function of competence and uncertainty.

14 Having proposed a model in which voters' "party difference in benefits" is assumed to be constant, Achen (1992, 200) specified that the model was intended to apply "within realignments" of the party system. With major realignments typically occurring every 30 or 40 years (Sundquist 1983), this suggests that a typical voter will have a handful of terms of experience on which to base his estimate of party performance. Allowing performance to vary over time makes the voter's prior experience less informative than in the simplest version of the model; the implications of randomness in that case are qualitatively similar, but more severe.

As the magnitude of random forces (represented on the horizontal axis in figure 4.1) shrinks to zero (at the left axis), the reelection probability for any incumbent with greater-than-average competence approaches its maximum value of 1.0.<sup>15</sup> However, as the proportion of luck in the admixture of competence and luck experienced by voters increases, the probability of reelecting a competent incumbent declines. The decline is especially rapid when the incumbent's relative competence is only modest, as in the lowest of the three lines in figure 4.1. These results underline the extent to which random retrospection is likely to impair voters' judgment.

In our model of retrospective sanctioning, the key issue is how much effort retrospective voters can induce incumbents to exert on their behalf. Figure 4.2 shows how this equilibrium level of effort—and thus the voter's expected well-being—varies with the extent to which electorally relevant subjective well-being depends upon luck rather than incumbent effort. The implication of the figure is that, in any system where "moral hazard" is a concern—that is, in any system in which incumbents face a trade-off between desiring reelection and other goals—random retrospection can significantly reduce incumbents' diligence, degrading the efficacy of elections as mechanisms for inducing accountability. Indeed, as the magnitude of random forces increases, the incumbent's equilibrium level of effort goes to zero (since no feasible exertion of effort can increase her chances of reelection by enough to be worthwhile).

In light of these theoretical patterns, one might expect scholars of retrospective voting to have devoted substantial effort to assessing the magnitude of randomness in the perceptions and behavior of voters and to exploring the bases of that randomness. However, that has not been the case. Although allusions to the potential difficulties faced by retrospective voters appear with some frequency, they are usually made in passing and are often remarkably brief and vague.

Downs, one of the first and most influential theorists of rational retrospective voting, clearly recognized the problem. He wrote (1957, 45–46),

we have glibly spoken of voters computing their party differentials and performance ratings without pointing out how difficult such com-

<sup>15</sup> Reelection probabilities for incumbents with negative (less than average) competence are a mirror image of those shown in figure 4.1. As the magnitude of random forces shrinks to zero, the probability of reelecting an inferior incumbent shrinks to zero; but increasing randomness increases the probability of reelecting an inferior incumbent, especially when her relative competence is only modestly negative.

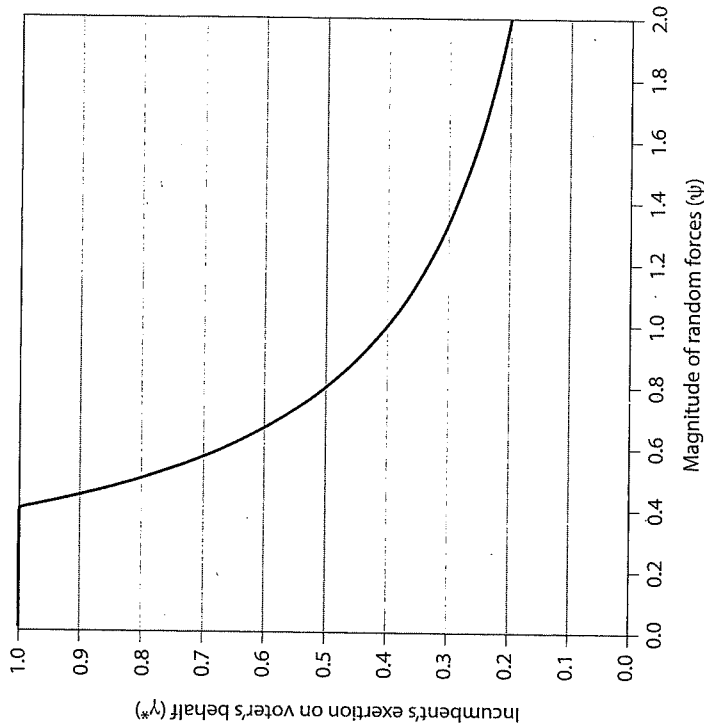


Figure 4.2. Blind Retrospection Erodes Incumbents' Diligence

putation is. In order to find his current party differential, a voter in a two-party system must do the following: (1) examine all phases of government action to find out where the two parties would behave differently, (2) discover how each difference would affect his utility income, and (3) aggregate the differences in utility and arrive at a net figure which shows by how much one party would be better than the other. This is how a rational voter would behave in a world of complete and costless information—the same world in which dwell the rational consumer and the rational producer of traditional economic theory.

In the real world, uncertainty and lack of information prevent even the most intelligent and well-informed voter from behaving in precisely

prove public understanding of environmental risks" (Science Advisory Board 1990).

Nor do voters seem to be very good at discerning the extent to which "thugs make neighborhoods unsafe." Public perceptions of the prevalence of crime seem to be more strongly related to local media coverage of crime—largely a product of broadcasters' commercial incentives—than to official crime rates (Giraber 1980). Thus, Gallup polls typically find majorities of respondents saying that crime is increasing, and pluralities saying that crime *in their area* is increasing, even though recent actual rates of violent crime as well as property crime have generally leveled off at extremely low numbers" (Saad 2007). Misperceptions of this sort are probably common in domains where voters have very limited personal experience, even (and perhaps especially) when, as in the case of crime, they care a lot about the true state of the world.

Even perceptions of economic well-being—a domain in which objective measures of social conditions are extraordinarily well-developed and salient—are subject to considerable vagaries. Prospective voters' economic perceptions are powerfully shaped by partisan biases, rationalization, and sheer randomness (Kramer 1983; Conover, Feldman, and Knight 1987; Bartels 2002a; Erikson 2004). Scholars who recognize that fact often assume nonetheless that *the electorate as a whole* responds sensibly and sensitively to actual economic experience under the incumbent administration. As James Stimson (2004, 165) put it, "The net perception of the economy, neither biased nor ignorant, is . . . right on the mark."

In our view, Stimson's reassuring formulation significantly exaggerates the reliability of economic retrospections. Aggregate economic perceptions are far from perfectly correlated with objective economic indicators.<sup>16</sup> And discrepancies between objective economic conditions and the public's economic mood can be politically consequential, as they probably were in 1992, when

16 In support of his claim, Stimson (2004) cited the analysis of Erikson, MacKuen, and Stimson (2002, 98) reporting a correlation of .54 between "Business Retrospections" (the economic assessment most similar to those employed in most studies of retrospective economic voting) and the previous year's income growth rate; but a correlation of .54 means that just under 30% of the variance in economic perceptions was explained by income growth, leaving plenty of room for bias and ignorance. Erikson, MacKuen, and Stimson's more elaborate regression analyses predicting aggregate economic expectations and perceived good or bad "economic news" on the basis of leading economic indicators, inflation, and unemployment produced stronger correlations, but the strongest predictors in those analyses were not objective economic indicators but lagged *perceptions* (2002, 94). That gives people credit for having the same bad judgments this period that they had last period; it is not evidence that voters are "right on the mark."

the fashion we have described. Since he cannot be certain what his present utility income from government is, or what it would be if an opposition party were in power, he can only make estimates of both. . . . When we open the door of our model to uncertainty, we must also admit such undesirables as errors, false information, and ignorance.

Having admitted the possibility of "errors, false information, and ignorance" in voting behavior, Downs had little more to say about their likely nature, magnitude, or consequences. In similar fashion, Key (1966, 110) acknowledged but then devoted little further attention to the importance of "the impressions of the march of affairs that exist in the minds of the voters. Differences in voters' interest, in their range of information, in the orientation of their attention, in their first-hand experience, and in their exposure to communications produce enormous variation in their perceptions of events and, consequently, in their appraisals of the alternatives posed by the electoral system."

Subsequent analysts have mostly followed suit, ignoring or dismissing substantial potential challenges to effective retrospective voting. Take, for example, Fiorina's (1981, 5) striking claim that retrospective voters can "ascertain whether the incumbents have performed poorly or well" simply by "calculat[ing] the changes in their own welfare." Of course, this claim is not literally true. If our voter slips on the ice and breaks his leg on the way to the polls, his own welfare has clearly changed for the worse, but we would not expect him to infer from his misfortune that the president has performed poorly. Fiorina's claim makes sense, however, if we assume that voters know a good deal about *which* "changes in their own welfare" are plausibly attributable in some part to the incumbents' performance. That is, they need to monitor indices of incumbent performance that contain a lot of honest signal and not too much noise. For some voters, domains of evaluation, and political contexts, that assumption may be a useful starting point; for others, not.

How are Fiorina's retrospective voters supposed to monitor the extent to which "polluters foul food, water, or air"? Few of the most serious threats to environmental well-being are immediately evident to the senses. Indeed, when the staff of the U.S. Environmental Protection Agency compared public perceptions of the relative significance of various environmental threats with those of experts, they found little correlation. The agency's Office of Policy Analysis (1987, 95) noted that public misperceptions of environmental threats seemed to be skewing the EPA's priorities, inspiring the EPA's Science Advisory Board to recommend that the agency "should work to im-



retrospective voting electorate and reaction anticipating politicians provides latitude for political leadership.<sup>17</sup>

The theories of retrospective voting we have considered assume that voters base their choices at the polls entirely on assessments of how much the incumbent party has contributed to their own or the nation's well-being. However, when voters have their own ideas about good policy, sensible or not, they may be tempted to vote for candidates who share those ideas, as in the spatial model of voting discussed in chapter 2. In that case incumbent politicians may face a dilemma: should they implement the policies voters want or the policies that will turn out to contribute to voters' welfare?

The strategic implications of this dilemma are nicely captured in the formal model of "leadership and pandering" developed by Brandice Canes-Wrone, Michael Herron, and Kenneth Shotts (2001). They posited a world in which an elected official has some private information regarding which of two alternative policies would serve the voters' interest. The elected official chooses a policy that may or may not comport with the preferences of voters, who do not have access to the elected official's private information.<sup>18</sup>

The impact of the elected officials' policy choice on voters' welfare may or may not be revealed before the next election. If it is, voters revise their beliefs regarding the elected official's quality accordingly, and an incumbent who is revealed to have chosen the right policy is more likely to be reelected than one who is revealed to have chosen the wrong policy.<sup>19</sup> However, if the impact of the policy is not revealed before the next election, voters may still condition their support on whether the incumbent chose the policy that they believe (but do not yet know) will better serve their interests.

If the incumbent's private information suggests that the policy the voters prefer will, in fact, maximize their welfare, then her choice is straightforward. However, if her private information suggests that the voters are mistaken, then she may have to choose between the risk of being punished for flaunting their will (if she chooses the policy she thinks is right, but its impact is not revealed before the next election) and the risk of being punished for choosing

17 The assumption that voters are relatively but not completely uninformed is crucial to the analysis. If voters knew as much as their leaders about the likely consequences of policy choices, there would be no need for leadership. On the other hand, if voters paid no attention to policy (as in the purely retrospective models we have outlined in this chapter), leaders would have no incentive to "pander."

18 The model assumes—unrealistically—that voters, having observed the impact of only one of the two alternative policies, will know whether it was the right choice or the wrong choice.

misleadingly negative media coverage of an improving economy contributed to widespread economic pessimism and the defeat of incumbent president George H. W. Bush (Hetherington 1996).

Moreover, even when aggregate economic perceptions do track real changes in economic conditions, they may provide a misleading basis for assessing voters' genuine well-being, and thus for selecting or sanctioning incumbents. For example, in chapter 6 we shall show that American voters focus inordinately on *recent* economic conditions, forgetting or ignoring most of their relevant experience under the incumbent and voting solely on the basis of how they feel about what has happened lately.

It does not seem unreasonable to imagine that a party's recent past performance might provide voters with a useful clue about its likely future performance. However, in light of the crucial importance of this assumption for the whole notion of retrospective selection, it is striking that it has never, as far as we know, been subjected to any systematic empirical examination. Most theorists of retrospective voting have simply *assumed* that there are real, persistent differences in competence between competing teams of political elites, and that voters (consciously or unconsciously) evaluate incumbents on the basis of criteria that do, in fact, systematically predict future performance.<sup>17</sup> In chapter 6, we put those assumptions to the test.

#### LEADERSHIP AND PANDERING

One of the most attractive features of theories of retrospective accountability is that they free elected officials from subservience to the policy preferences of ordinary citizens—preferences that are often likely to be vague, uninformed, or incoherent. As Fiorina (1981, 201) put it, retrospective voting "lays no policy constraint on the governing administration; rather, the government is free to innovate, knowing that it will be judged on the results of its actions rather than their specifics. In a word, the accountability generated by a

17 We note in passing that the entire notion of retrospective accountability also hinges critically on the assumption that there are, in fact, competing teams of political elites willing and able to seek reelection. In his classic study of southern politics in the Jim Crow era, Key (1949, 302–305) observed that an absence of party competition, stringent term limits, and disorganized "kaleidoscopic" factional competition within the dominant Democratic Party often left voters with no one to reward or punish: "The candidates are new and, in fact, deny any identification with any preceding administration. Without continuing groups, there can be no debate between the 'ins' and 'outs' on the record." The resulting "aromized and individualistic politics," he argued, frustrated political accountability while placing "a high premium on demagogic qualities of personality that attract voter-attention."



the wrong policy (if she chooses the policy they think is right, but the negative consequences become evident before the next election).<sup>20</sup>

Canes-Wrone, Herron, and Shortt's analysis identifies the circumstances under which an incumbent may be tempted to "pander"—implementing the policy preferred by the voters even though her private information suggests that it will not serve their interests.<sup>21</sup> In their basic model, pandering occurs when the probability is not too high that the inferiority of the voters' preferred policy will be revealed before the next election (and if the expected quality of a prospective challenger is strong enough that the incumbent has to worry about her public standing at election time).<sup>22</sup>

Alexander Hamilton argued in Federalist Number 71 that politicians' temptation to pander depends in part on the length of their terms in office, with longer terms encouraging politicians "to be the guardians of those [gen- uine] interests to withstand the temporary delusion." Canes-Wrone, Herron, and Shortt's analysis has the same implication: the more distant in time is the next election, the more likely bad policies are to be revealed as such by the time the voters go to the polls. An intuitive recognition of this fact may help to account for the fact that American governors' terms have been steadily lengthened since the late 18th century, when one-year terms were most common.

For lower-level offices, however, a good deal of variation in term lengths remains, and it seems to have just the sort of consequences suggested by Hamilton and by Canes-Wrone, Herron, and Shortt's analysis. For example, elected officials facing the issue of fluoridating drinking water in the 1950s and 1960s were significantly less likely to pander to their constituents' ungrounded fears when longer terms gave them some protection from the "sudden breezes of passion" that Hamilton associated with public opinion. Figure 4.3 shows the dramatic difference that longer terms made to mayoral

20 In a detailed study of politics and policy-making in Latin America, Stokes (2001) undefined the tension between populist campaign promises and "neoliberalism by surprise" once the winners took office.

21 Technically, elected officials in Canes-Wrone, Shortt, and Herron's model are of two types. "High-quality" incumbents know with certainty which policy will maximize voters' welfare, and it is never in their electoral interest to act contrary to their own judgment. The dilemma of whether to "lead" or "pander" arises for "low-quality" incumbents, whose private signals are informative but not perfect.

22 Canes-Wrone, Herron, and Shortt also consider a more complicated model in which some policy options are more likely than others to be revealed as good or bad before the next election. In that case, additional strategic possibilities arise, including such perverse strategies as "fake leadership"—choosing a policy that is both unpopular and (probably) inferior in hopes of convincing voters that the incumbent must know best.

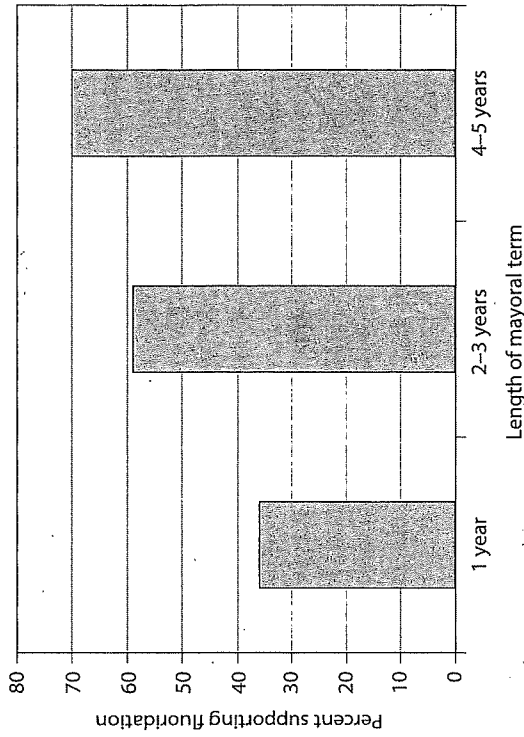


Figure 4.3. American Mayoral Terms and Support for Fluoridation

support for fluoridation.<sup>23</sup> Many political leaders, not caring deeply about the topic, ducked; but those with longer terms had more political leeway to do what was right, and a significant fraction of them used it.

#### RATIONAL CHOICE OR KICKING THE DOG?

So far, we have examined the logic of retrospective voting on the *assumption* that voters, however uninformed they may be, are at least behaving rationally. As Achen (1992, 199) put it, following Downs (1957, 106): "Retrospective voters' . . . choose at least partly on the basis of the past, even though they cannot hope to change it and its lessons may have long ceased to be relevant. For the rational voter, however, bygones are bygones, and only the future matters. . . . As standard decision theory teaches, the rational chooser looks forward, not backward. The rational voter is a prospective voter, and the past is useful only for its clues about the future." By this logic, voters may use the past to learn about the incumbent's likely future behavior, or they may reward

23 The data were reported by Crain, Kazz, and Rosenthal (1969, 179-183). The percentages shown in the figure are based on 11, 75, and 54 cases, respectively.

or punish the incumbent for something that has already happened, but only as a means of enforcing discipline, maintaining incentives for future incumbents to attend to their interests. In the oft-quoted words of V. O. Key, Jr. (1964, 568), the electorate is "a rational god of vengeance and of reward."

Scholars who have quoted Key's colorful phrase have mostly failed to note that he used it derisively. "The Founding Fathers," he wrote in the final edition of his influential textbook on party politics, "by the provision for midterm elections, built into the constitutional system a procedure whose strange consequences lack explanation in any theory that personifies the electorate as a rational god of vengeance and of reward."

In the first edition of the same textbook, Key (1942, 628) offered an even clearer dismissal of the rational interpretation of retrospective voting, noting that voters seem to have rewarded and punished incumbents at the polls for good or bad times

even before it could be said that the national Government could do much of anything to improve their condition. . . . If the party control of the national Government had little or nothing to do with their fate, how is this behavior to be explained? Is it to be considered as a rational seeking to better one's status by the ballot or is it merely blindly striking a blow at a scapegoat? To throw out the "ins" probably had about the same effect on economic conditions as evangelical castigation of Satan has on the moral situation. Perhaps the swing against the "ins" can best be described as a displacement of economic resentment on political objects. By this catharsis discontent was dissipated and the peace kept.

Readers whose only exposure to Key's work is to *The Responsible Electorate*, or (especially) to simplistic summaries of that book, will be surprised to find Key characterizing retrospective voting in such bleak terms.<sup>24</sup> However, the apparent irrationality of the electorate's vengeance and reward has been remarked by many political observers over the years. Indeed, in both of the 19th-century examples we cited earlier, the prominent politician on the losing end of the electorate's vengeance wrote feelingly of its illogic. President Buchanan, whose party was stung by the effect of the Panic of 1857, protested

<sup>24</sup> This passage disappeared from the second (1947) and subsequent editions of Key's textbook, replaced by more detailed analysis of New Deal voting patterns. It is unclear whether this alteration reflects a change in Key's view or merely routine updating, of which there was a good deal from one edition of Key's text to the next.

that "the administration are as responsible for the motions of the Comer as for the low price of iron" (Huston 1987, 167). For his part, Prime Minister Disraeli complained (Monyenny and Buckle 1929, 1395), "Never was so great a discomfiture with a cause so inadequate. I think, as far as I can collect, 'hard times' was the cry against us. The suffering want a change—no matter what, they are sick of waiting." Presumably, voters in Victorian Britain did not imagine that their government controlled the weather, but that did not prevent them from exacting vengeance at the polls in the wake of a wrenching succession of bad harvests.

In the post-Keynesian era, governments are thought to have real influence on the course of their national economies. Economists tell us that competent policy-makers can reduce the likelihood of recessions and ameliorate their effects when they occur; and practicing politicians take their assessments and advice seriously. Thus, the strong tendency for voters to reward incumbents for good economic times and punish them for bad times is readily interpretable as reflecting rational retrospective voting along either or both of the lines we have sketched in this chapter.

However, the fact that voters rewarded and punished incumbents for economic booms and busts well before Keynes taught governments how to exercise (at least partial) control over their economies represents a significant embarrassment to the rational interpretation of retrospective voting. Moreover, the rational basis for economic voting even in modern industrial economies is murky at best, especially in small countries highly vulnerable to global economic tides. The economy may flourish or falter, but political leaders may or may not be responsible. Moreover, whatever they did right or wrong may or may not be relevant in the circumstances, possibly quite different, that prevail after the election.

Does it really matter whether retrospective voting is thought of as "a rational seeking to better one's status by the ballot" or "merely blindly striking a blow at a scapegoat"? Debating whether a given pattern of behavior is or is not "rational" is largely a tedious matter of definition. Largely, but not entirely. Insofar as we think of retrospective voting as reflecting conscious, conscientious efforts by voters to further their future individual and collective well-being, we should expect them to overcome—as best they can, given inevitable constraints on their attention and information—the challenges to effective retrospective surveying in the preceding section. However, if retrospective voting represents "a displacement of economic resentment on political objects," as Key put it, then there is no particular reason to expect that displacement to be bound by careful thinking about cause and effect.

Some previous studies have attempted to gauge the “rationality” of retrospective voting by focusing specifically on the electoral impact of forces beyond incumbents’ control. For example, Justin Wolfers (2002) showed that U.S. governors in oil-producing states were rewarded and punished by voters for economic fluctuations traceable to changes in world oil prices. He concluded that the “simplest explanation” of his results was that “voters are quasi-rational” and “make systematic attribution errors” (Wolfers 2002, 16–17). Similarly, Daniela Campello and Cesar Zucco (2013) found that the fortunes of political leaders in Latin America could be predicted on the basis of international commodity prices and U.S. interest rates, a result that they suggested “should prompt democracy enthusiasts to engage in some soul-searching.” And Bartels (2014) found that incumbent parties in a variety of affluent democracies were punished at the polls for declines in their national economies in the wake of the Great Recession, regardless of how they fared relative to other OECD economies at the time.

While these studies and others in the same vein shed useful light on electoral behavior and political accountability, they shed little light on the underlying psychology of retrospection. Were voters swayed by world oil price shocks doing “a reasonable, albeit imperfect job of disentangling variation in the economy due to competence from variation due to other factors” (Wolfers 2002, 17), or were they punishing their leaders out of mere frustration?

Rather than quibble about which specific economic conditions can or cannot sensibly be attributed to incumbent politicians, in chapter 5 we focus on events that even “reasonable, albeit imperfect” rational voters ought to recognize as clearly outside any incumbent’s control—droughts, floods, and shark attacks. We argue that voters respond to those events in much the same way that they respond to ordinary economic downturns, and thus that most retrospective voting of all kinds is more a matter of kicking the dog than of rationally assessing blame or credit. However, that debate makes little difference for democratic theory. Even if the voters are doing the best they rationally can manage, the result of their efforts is sadly unimpressive, as we shall show. Hence, no matter how we interpret voters’ responses to bad times, the “blindness” of retrospection poses a significant challenge to democratic accountability.

## CONCLUSION

In this chapter we have examined a stylized world in which ordinary voters think quite differently about politics than they are supposed to do in the conventional spatial model of voting outlined in chapter 2. These “retrospective”

voters have no policy preferences representable as “ideal points” in multidimensional space. They may know little or nothing about the policies any government has actually enacted. They only know whether things have gone well or badly under the incumbents, and they vote on the basis of that retrospective evaluation. As Key (1966, 7) insisted, such “voters are not fools”—but they are woefully ignorant about much of what is typically taken to be at the heart of democratic politics.

The theory of retrospective accountability demonstrates that, in principle, voters in this stylized world can select competent leaders and discipline those leaders to pursue the voters’ well-being, even though they cannot distinguish “politically relevant” pain and pleasure—aspects of well-being plausibly attributable to the policies of the incumbent government—from everything else going on in their lives. However, the *effectiveness* of retrospective accountability depends significantly on the precision with which voters can discern the contributions of the incumbent government to their well-being. Indeed, the implication of the formal models set out in this chapter and summarized in figures 4.1 and 4.2 is that even relatively small amounts of randomness in the relationship between incumbents’ actions and voters’ subjective well-being can significantly degrade the efficacy of elections as mechanisms for selecting and sanctioning political leaders.

The theory of retrospective voting promises a realistic account of democratic politics. But like any theory, it is a proposal, not a settled finding. How well can the voters manage what the theory of retrospective voting asks them to do? In the next three chapters we provide a more detailed assessment of how, and how well, retrospective voting works in actual elections.

reign, and perhaps his life as well (Bell 1971; Hassan 1994). Not surprisingly, there are records of Egyptian court officials wishing their pharaoh a good Nile flood.

Through the centuries, rulers and their potential challengers have been well aware of the potential political significance of natural disasters. Poor weather and bad harvests have been given substantial credit for the rise of the Populists in Kansas (Miller 1925) and Nebraska (Dixon 1898, 637; Barnhart 1925) in the 1890s. More recently, one Republican presidential campaign official said in 1992 that "it wouldn't be so bad" if Hurricane Andrew left Florida and instead "blew on up to Kentucky and the rust-belt states" where incumbent George H. W. Bush had less chance to win electoral votes (Schneider 1995, 96).

When disasters take on truly catastrophic dimensions, not just the ruler but the entire regime may face a crisis of legitimacy. Islamic missionaries in Java and Sumatra successfully blamed Dutch rule for the 1883 volcanic explosion on Krakatoa (Winchester 2003, 317–338). An earlier catastrophic eruption in 11th-century Arizona apparently triggered social upheavals among the people living nearby; their Hopi descendants still preserve folk memories of the event, which they interpret as retribution for their ancestors' "morally imbalanced and corrupt" culture (Gidwitz 2004, 52). Similarly, the European famine of the early 14th century led to outbreaks of heresy and heretic burning in Silesia (Jordan 1996). Later in the 14th century the Black Death, which may have killed one-third or more of the population of Europe, generated numerous spontaneous religious and political movements to threaten church and government (Herlihy 1997, 64): "The plague also discredited the leaders of society, its governors, priests, and intellectuals, and the laws and theories supported by them. These elites were obviously failing in their prime social function, the defense of the common welfare, in the name of which they enjoyed their privileges."

The argument that natural disasters threaten rulers and regimes is not new. However, the base of evidence on which it rests, while impressively broad historically, is also uncomfortably thin. In this chapter we take up the challenge of providing more precise and comprehensive evidence about the political impact of natural disasters. We focus on modern history and electoral politics, where the data are sufficiently detailed and reliable to trace the political consequences of hard times.

Contemporary democratic rulers have little aura of divinity about them, nor have they faced epic famines or medieval plagues. Nonetheless, we find

## CHAPTER FIVE

# Blind Retrospection: Electoral Responses to Droughts, Floods, and Shark Attacks

And Moses stretched forth his rod over the land of Egypt, and the Lord brought an east wind upon the land all that day, and all that night; and when it was morning, the east wind brought the locusts. And the locusts went up over all the land of Egypt, and rested in all the coasts of Egypt: very grievous were they; before them there were no such locusts as they, neither after them shall be such. For they covered the face of the whole earth, so that the land was darkened; and they did eat every herb of the land, and all the fruit of the trees which the hail had left: and there remained not any green thing in the trees, or in the herbs of the field, through all the land of Egypt.

Then Pharaoh called for Moses and Aaron in haste; and he said, I have sinned against the Lord your God, and against you.

—Exodus 10:13–16 (King James Version)

When collective misfortune strikes a society, someone must be blamed. For ancient Israel, disasters were God's punishment for sin—perhaps the ruler's sin, perhaps the people's. Theology did not single out the guilty party, but it structured the search and set limits on what counted as a credible explanation.

In the theology of classical Egypt, pharaohs were divine beings responsible for making the Nile flood annually. Some scholars believe that when it failed to do so, as happened repeatedly in the First Intermediate Period (ca. 2200 BCE), the resulting famines and political disorder shortened the pharaoh's

that when election time comes, the electorate continues to hold rulers responsible for calamities and disasters that are clearly beyond their control. We go beyond that simple correlation, however, to argue a deeper point. While voters' reactions to natural disasters are of importance in their own right, our interest here is in what they can tell us about democratic accountability more broadly. From that perspective, electoral responses to natural disasters are just particularly illuminating instances of the broader phenomenon of retrospective voting.

Our assertion is that voters' retrospections are blind, not just in natural disasters but in hardships of all kinds. When they are in pain they are likely to kick the government, so long as they can justify doing so with whatever plausible cultural constructions are available to them. Only if no such constructions are available, or if no ambitious challengers emerge to articulate them, will people take out their frustrations on other scapegoats, or just suffer. In most cases, incumbents will pay at the polls for bad times, whether or not objective observers can find a rational basis for blame.

Our analysis begins with the unjustly neglected electoral impact of shark attacks.

#### SHARK ATTACKS IN NEW JERSEY: THE VOTERS BITE BACK

On the four-day Fourth of July weekend in 1916, the beaches of New Jersey were packed with crowds happy to escape the summer heat of nearby cities.<sup>1</sup> On Saturday, July 1, a young Ivy League graduate from Philadelphia, Charles Vansant, was swimming just beyond the breakers in four feet of water at Beach Haven when he was attacked by a shark. Skillful lifeguards managed to get him to shore, but he died soon after from blood loss. Five days later, a young Swiss bellhop named Charles Bruder, a strong swimmer like Vansant, ventured out past the lifelines at Spring Lake beach, some 45 miles north of Beach Haven. He, too, was attacked by a shark. Though rescued by lifeguards in a small boat, he died of his wounds before reaching shore.

In the days after the two deaths, nearly all of the diminished numbers of Jersey Shore swimmers stuck close to shore. However, no one worried about boys swimming in a creek on July 12 in the town of Matawan, about two miles from open water. Yet one was attacked and killed by a shark, as was

1 Unless otherwise noted, our historical account follows that of Fernicola (2001); we also draw upon Capuzzo (2001).

a young man from the town who dove in to recover the boy's body. Downstream, another group of boys were swimming at the same time in ignorance of the attacks. Within half an hour, one of them had his leg mauled by a passing shark. However, he was quickly pulled from the water, reached the local hospital, and survived.

By this time, the mounting panic reached a crescendo. Even the distant *San Francisco Chronicle* trumpeted the shark attacks in a July 14 front-page headline: "EAST COAST BEGINS WAR ON RAVENOUS MAN-EATERS" (Fernicola 2001, 87). Steel mesh was being installed at beaches. Bounties were offered, and sharks were killed in sizable numbers along the shore. Finally, one great white shark was hauled in near Matawan Creek with what appeared to be human bones in its stomach. Perhaps for that reason, the attacks stopped, ending the most serious string of shark-related fatalities in American history.

Before the attacks, no arm of government had patrolled for sharks or set up barriers against them in New Jersey, since there had never been a recorded shark attack in the history of the state. Indeed, prominent American scientists doubted that unprovoked shark attacks on human beings ever occurred, certainly not as far north as New Jersey (Fernicola 2001, 22).<sup>2</sup> The general climate of skepticism led the *New York Times* to bury its article about the first attack on page 18, headlined "Dies after Attack by Fish"—no doubt a consolation to the New Jersey resort owners, who were anxious to avoid publicity.<sup>3</sup>

In the aftermath of the attacks, the federal government was called on for help. The resorts were losing money rapidly, with \$250,000 in reservations cancelled within a week. Some resorts had 75% vacancy rates in the midst of their high season (Capuzzo 2001, 274). Losses may have amounted to as much as \$1 million for the season altogether, a sizable sum in 1918 (Fernicola 2001, 174). Letters poured into congressional offices from the affected communities demanding federal action, though there was little any government agency could do. Fernicola (2001, 70) described the atmosphere, as the shark attacks entered popular imagery and became a metaphor for other political crises as well: "Newspaper cartoons now portrayed Wilson's chances for reelection in November, using the shark fin as the symbol for his potential loss.

2 Indeed, two scientists who were later called in to investigate the attacks, Dr. John T. Nicols, an ichthyologist and director of the Fishes Wing of the American Museum of Natural History, and Dr. Frederick Lucas, director of the museum, had recently coauthored with a third scientist an article arguing that unprovoked sharks never attack human beings.

3 Parallels to the film *Jaws* and its sequel are no accident. Peter Benchley, the author of the novel on which the film was based, was a New Jersey resident, and the film version, though set on Long Island, New York, included a reference to the 1916 New Jersey attacks.



The black fin labeled 'defeat' was shown slicing through shark-infested northeast regions. Other political cartoons of the day showed lawyers, represented by sharks heading toward a beleaguered sailboat, embossed with 'Union Bank.' At the stern of the bank boat, a chewed and legless victim dangled over the gunnel depicting 'deposits.'"

As it happened, the Secretary of the Treasury, William McAdoo, had a summer home in Spring Lake and was in residence at the time of the second attack. Joseph Tumulty, Wilson's powerful aide for political affairs, had a summer home in Asbury Park, about five miles north of Spring Lake. President Wilson himself, a former president of Princeton University and former governor of New Jersey, had been looking for a summer White House in New Jersey as well, and chose a hotel in Asbury Park, moving there shortly after the attacks ended. Thus the attacks received immediate federal attention. Wilson held a cabinet meeting to discuss the attacks (Fornicola 2001, 70), but the Bureau of Fisheries could suggest nothing beyond killing sharks at random and warning bathers. "No certainly effective preventive measure could be recommended," they said (Capuzzo 2001, 277). The president could only direct the Coast Guard to inspect the beaches and patrol the water. However, the shark attacks ended and autumn arrived before much could be done.

By election time in November, Wilson was back at his Asbury Park headquarters, but other election issues, notably potential U.S. entry into World War I, took over the headlines (Link 1954, 247-251). In the end, Wilson lost nearly all the northeastern and Great Lakes states, including New Jersey, but managed to squeak out his reelection by adding most of the Great Plains, Mountain States, and West to the Democrats' customary Solid South.

Did the shark attacks influence the presidential election in the affected areas of New Jersey? Hitherto, sharks have not been suspects in any electoral analysis. Nonetheless, if our argument is correct, they should have reduced Wilson's vote. First, the attacks caused several deaths plus considerable emotional and financial distress to shore communities. Second, the election occurred just a few months after the summer's events, increasing the likelihood that they would be fresh in the minds of the voters as they went to the polls. Third, high federal officials were present at the scene from the beginning, reinforcing the notion that the federal government should have done *something* to deal with the crisis. The fact that no government has any influence over sharks would, from our perspective, have been irrelevant to the voters.

The evidence for a shark effect turns out to be rather strong. We now turn to the first piece of that evidence, using election returns from New Jersey

counties.<sup>4</sup> The Wilson vote in 1916 is the outcome to be explained. Our key explanatory factor is an indicator for "beach counties," defined as Monmouth, Ocean, Atlantic, and Cape May counties. These were, and are, the classic "Jersey Shore" counties listed in the guidebooks, whose beach areas are heavily dependent upon summer tourism. They are the places in which the shark attacks would have had the most pronounced economic effects. The attacks themselves took place in Monmouth (three deaths) and Ocean (one).

We include two additional factors in our country-level analysis. The first is the Wilson vote in 1912, a measure of both partisanship and candidate appeal, including favorite son effects. Wilson's 1912 vote predicts his 1916 showing well, despite the fact that 1912 was a three-way race with former president Teddy Roosevelt running as a Progressive.<sup>5</sup> By contrast, the four presidential elections prior to 1912 (and their mean) were less correlated with the 1916 vote, and they added nothing to the accuracy of the statistical analysis once 1912 was included.<sup>6</sup>

One other control variable is needed to capture an important change in New Jersey politics between 1912 and 1916. Having supported Wilson for governor in 1910, the New Jersey bosses turned against him shortly after his election.<sup>7</sup> They initially opposed his nomination for president in 1912, but fell in line once it became inevitable (Link 1947, chaps. 8-9 and 427-428). After he became president, however, Wilson's control of the New Jersey Democratic Party, once nearly complete, slipped away (Blum 1951, 39, 76; Link 1947, 288). For example, the infamous Jersey City political boss Frank Hague supported a progressive Wilson ally during this period (McKean 1940, chap. 3; Connors 1971, chap. 3). To take account of Wilson's reduced power over the

4 New Jersey electoral data are from the official reports published in the *Legislative Manual of the State of New Jersey*, various years.

5 Throughout the Northeast, the Roosevelt vote from 1912 returned almost entirely to Charles Evans Hughes, the Republican candidate, in 1916. (Socialist and other minor candidates, including Prohibition advocates, were also running in both years, but of course only Roosevelt was a serious third-party contender for the presidency.) Wilson gained less than a percentage point statewide in New Jersey in 1916 from his 1912 totals, and similar results held in other northeastern states. Wilson's 1912 vote is an excellent predictor of his 1916 vote across New Jersey counties, and even at the township level.

6 Adding the Roosevelt proportion of the vote from 1912 generated a small positive, statistically insignificant coefficient. Keeping the Roosevelt variable made no difference in subsequent analyses, and so it was dropped.

7 For this reason, Wilson's vote for governor in 1910 is poorly correlated with his showing in both presidential elections and was not used as a statistical control in our analysis of his 1916 vote.



bosses in 1916, we include a control variable for "machine counties," defined as those counties with at least 30,000 voters in 1916 and 60% or more "foreign" citizens in the census of 1910.<sup>8</sup> The counties so defined are Bergen, Hudson, Essex, and Union, adjacent to each other and just across the state line from New York City.

Two of these machine counties, Hudson (Jersey City) and Essex (Newark), were particularly well known for boss control. In fact, alone among New Jersey's counties, Wilson never did get so much as partial control of the Essex Democratic machine, which was under the thumb of James Smith, Wilson's bitter political enemy; throughout this period (Blum 1951, 39-40; Link 1947, 288, 424). For that reason, Wilson's 1912 vote in Essex was so low relative to its electoral history that the county becomes a substantial outlier in predicting the 1916 vote, even beyond its status as a machine county. Simply put, Essex County in this electoral period does not act like the rest of New Jersey at the polls; we therefore excluded it from our analysis. The other 20 New Jersey counties make up our sample.

Table 5.1 presents the results of a statistical analysis estimating the difference in Wilson's 1916 presidential vote share between beach counties and non-beach counties, controlling for machine counties and for Wilson's 1912 vote share. All of the parameter estimates are substantively significant and sensibly sized, and each of them is statistically significant beyond the .01 level. The analysis accounts for Wilson's 1916 vote share with an average error of just 1.7 percentage points, and the correlation between actual and predicted 1916 vote shares is .97.<sup>9</sup>

The estimated negative effect on Wilson's vote in the beach counties is a little more than three percentage points, with a 95% confidence interval confined between 1.3 and 5.2. The shark attacks indeed seem to have had an impact. The statistical significance of the estimate is due to the very consistent effect across the beach counties, as may be seen from figure 5.1. This figure shows the statistical relationship between Wilson's 1916 vote share and his 1912 vote share with the machine county variable controlled. The linear relationships are estimated separately for beach and non-beach counties, with Essex excluded. As the graph shows, the beach counties are each depressed

8 "Foreign" here means that the citizen was foreign-born or had at least one foreign-born parent (the so-called hyphens in the vernacular of the time).

9 None of the residuals from this regression analysis falls more than two standard deviations from zero, and only one of them (Salem's) is near that level, about what would be expected by chance. By contrast, the excluded Essex County observation has a residual 4.6 standard deviations from zero in this analysis, amply justifying its exclusion from the sample.

Blind Retrospection

Table 5.1. The Effect of Shark Attacks on the 1916 New Jersey Presidential Vote

Beach county	-3.2 (1.0)
Machine county	-5.7 (1.1)
Wilson 1912 vote (three-way fraction)	0.95 (0.06)
Intercept	4.5 (2.8)
Standard error of regression	1.7
Adjusted R <sup>2</sup>	.94
N	20

Parameter estimates from ordinary least squares regression analysis (with standard errors in parentheses) of Woodrow Wilson's vote share (two-party %) in New Jersey counties, 1916.

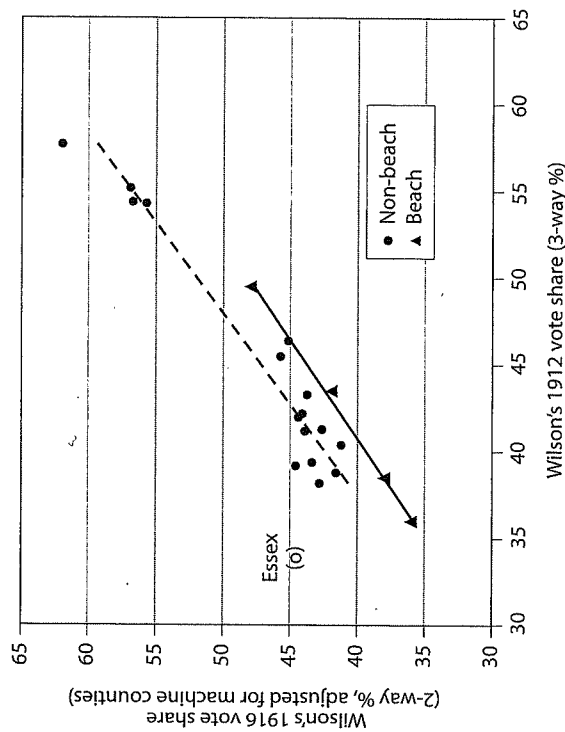


Figure 5.1. Support for Woodrow Wilson in New Jersey Counties, 1916 versus 1912

nearly the same amount from their expected 1916 vote, and the consistency of the effect bolsters the plausibility of the specification visually while tightening the standard errors statistically.<sup>10</sup>

We explored a variety of other statistical specifications using different measures of partisanship. None worked as well as Wilson's 1912 vote share, and the estimated effect of the shark attacks remained fairly constant—two to four percentage points—so long as the 1912 vote share was included. We also tried including measures of the proportion German, the proportion Irish, and the total proportion “foreign,” since speculation was rampant at the time of the 1916 election that these communities would be unhappy with Wilson over his potential entry into World War I on the British side or over the British suppression of the 1916 Easter Rising in Dublin. We found no effects of this sort, which is consistent with the conclusions of analysts of the national vote at the time.<sup>11</sup> Similarly, weighting the counties by their total 1916 turnout, or using the change in Wilson's vote share from 1912 to 1916 as the dependent variable, or both, never reduced the estimated impact of the sharks by more than a tenth of a percentage point. In fact, all the turnout-weighted estimates were larger by a few tenths of a percentage point. Thus the shark effect stands up well under a variety of alternative statistical specifications.<sup>12</sup>

We also undertook two additional investigations with different samples. First, we examined the vote in the first two shore townships where the attacks took place.<sup>13</sup> Both Beach Haven and Spring Lake were small, stable commu-

10 If the machine counties were counted as beach counties, they would fit nicely on the regression line. They are themselves on the water or adjacent to the Hudson River. Capuzzo (2001, 270–273) noted that fear extended well beyond the Jersey Shore counties, up through the machine counties and onto New York State beaches, where the economy was also harmed. One shark was killed with a revolver near a yacht club in machine-controlled Hudson County (Ferticola 2001, 27). Thus it is possible that some of the negative “machine county” effect is, in fact, due to the sharks.

11 Two days after the election (November 9), the *New York Times* headlined, “Both Candidates Got Hyphen Vote.” For subsequent treatments reaching the same conclusion, see Link (1954, 232–251) on the Germans and Leary (1967) and Cuddy (1969) on the Irish.

12 Another possibility we considered was that Roosevelt might have run worse in the beach counties than in the rest of the state, leaving Wilson fewer voters there to pick up from Progressive Republicans in 1916. This would have created an artificial drop in Wilson's 1916 vote in the beach counties. To the contrary, however, Roosevelt ran *better* along the shore than in the rest of the state, so that the shark attack effect is, if anything, slightly underestimated in table 5.1.

13 Marawan Township and Marawan Borough, where the final two shark deaths occurred in a river, were excluded from this analysis since they are not beach resort communities and thus suffered no widespread economic loss from their shark attacks or anyone else's. In any

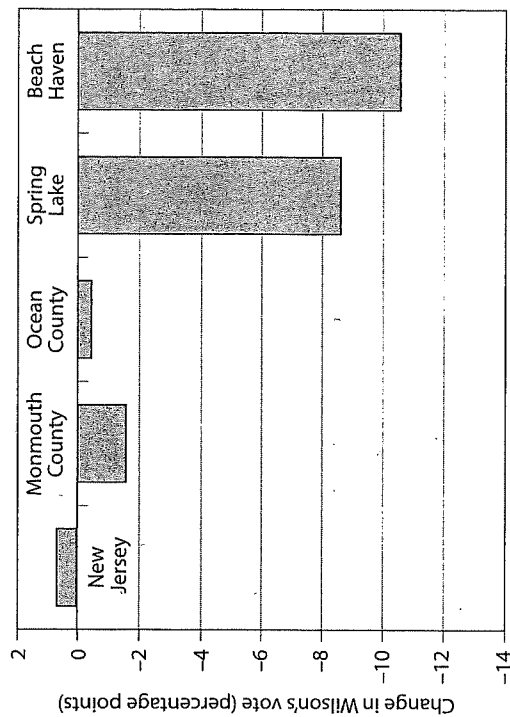


Figure 5.2. Change in Woodrow Wilson's Vote Share (1912–1916) in Counties and Townships with Shark Attacks

nities, making comparison sensible.<sup>14</sup> Figure 5.2 shows the vote change for Wilson between 1912 and 1916 in these two communities, and compares it with the change in their respective counties and in New Jersey as a whole. Both townships show remarkable drops in Wilson's support, 11 points in Beach Haven and 9 in Spring Lake—far more than the modest decline in the Wilson vote in their counties. It is apparent that something drastically reduced enthusiasm for Woodrow Wilson in these two townships.

We also investigated whether Beach Haven and Spring Lake were typical of beach areas. To answer this question, we examined the townships in Ocean County near the water. Ocean was chosen because it has many beach

case, the rapid growth in the number of voters in both places between 1912 and 1916 makes reliable comparison impossible; more than a quarter of the 1916 voters in Marawan Township had not been there in 1912.

<sup>14</sup> Beach Haven cast 112 votes for president in 1912 and 119 in 1916. The corresponding numbers for Spring Lake are 271 and 265.

communities, nearly all on a bank of land (Long Beach Island) clearly separated from the mainland. Thus there is no difficulty in separating those eight communities right on the beach, whose economies were damaged by the shark attacks, from the eleven towns near the beach but not on it, whose economies were less susceptible to harm.<sup>15</sup> New Jersey was growing rapidly in this era; to ensure that vote shares in 1912 and 1916 would be meaningfully comparable, we dropped townships whose vote totals grew or shrank by more than 25% in this four-year period.<sup>16</sup> This left us with 14 towns, 4 of them on the beach and 10 nearby. These two sets of communities had very similar overall Democratic percentages for Wilson in 1912 (37.1% at the beach and 33.5% in the near-beach), making them comparable.

In each area, we compared Wilson's vote percentages in 1912 and 1916. If our argument is correct, the beach voters should show the largest drop in support for Wilson, while the near-beach voters should have been largely unaffected. As shown in figure 5.3, the actual vote change turns out to be a drop of 13.3 percentage points in the beach area, compared to a tiny loss of half a percentage point in the near-beach area, an easily statistically significant difference.<sup>17</sup> Again, we find that disaffection for Wilson was widespread in the beach areas where livelihoods were most directly affected by the shark attacks, far different from the otherwise comparable areas next door, where Wilson's vote was nearly constant.

In summary, then, every indication in the New Jersey election returns is that the horrifying shark attacks during the summer of 1916 reduced

<sup>15</sup> The western border of the near-beach area was set to the current Garden State Parkway, which runs within a few miles of the shore in Ocean County.

<sup>16</sup> One beach township, Sea Side Park, apparently split into two between 1912 and 1916 and jointly nearly doubled in size; we dropped it from the analysis.

<sup>17</sup> The simplest approach is a differences-in-differences regression model weighted by the 1916 total vote (to take account of the wide range of electorate sizes in these boroughs). Thus, with the change in the Democratic vote percentage from 1912 to 1916 as the dependent variable and beach township as the explanatory variable, we obtain a coefficient of  $-12.8$  percentage points, with a standard error of 4.4 (and a  $t$ -statistic of 2.9). Alternately, a weighted regression with the Democratic vote in 1916 as the dependent variable, and with beach township and the 1912 Democratic vote as explanatory variables, yields a beach effect of  $-11.1$  percentage points with a standard error of 3.2 (and a  $t$ -statistic of 3.5). Unweighted regressions, though arguably substantively inappropriate, yield even larger beach coefficients. Finally, if we eliminate two townships with fewer than 50 voters in 1916, the differenced regression produces a beach coefficient of  $-8.4$  percentage points, while the second regression version yields  $-8.8$ , both with  $t$ -statistics exceeding 2.5. In short, alternate versions of the beach versus non-beach comparison lead to precisely the same substantive conclusion, which we summarize as a loss of about 10 percentage points in the areas most directly affected by the shark attacks.

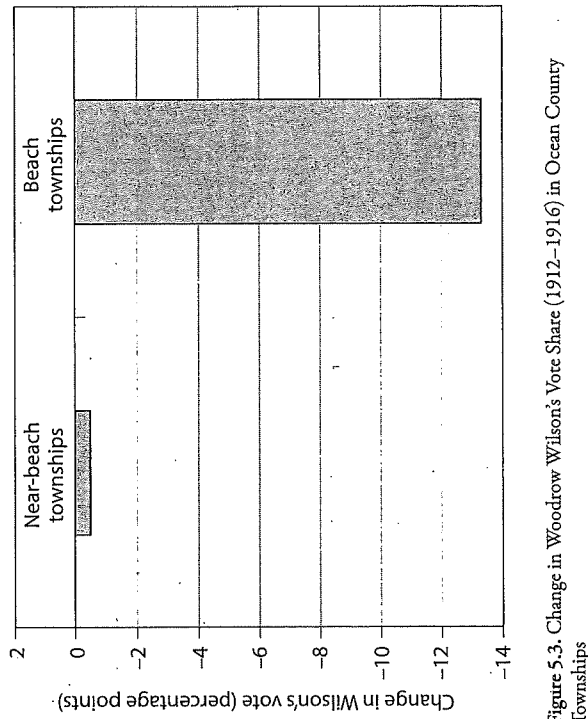


Figure 5.3. Change in Woodrow Wilson's Vote Share (1912–1916) in Ocean County Townships

Wilson's vote in the beach communities by about ten percentage points. An effect of that size may sound modest to those unfamiliar with American electoral experience, but by those standards it is a near-earthquake. (A full earthquake, the Great Depression, reduced Herbert Hoover's vote in New Jersey by 12 percentage points, from 59.8% in 1928 to 47.6% in 1932.)

In the case of the shark attacks, retrospection was surely blind. Shark attacks are random events in the purest sense of the term, and they have no governmental solution. If bathers insist on swimming in the ocean, governments then and now cannot save them, as subsequent attacks in New Jersey in 1960 and regular encounters in Florida, California, South Africa, and Australia demonstrate (Fericola 2001, chap. 5). Nor could the aftermath of the 1916 attacks be repaired by governmental action. The truth could not be covered up. The vacationers could not be compelled to come to the beach, nor could the sharks be forced to stay away. The government was helpless. Yet the voters punished anyway.<sup>18</sup>

<sup>18</sup> On one occasion, sharks apparently had a more direct and unfortunate impact on an incumbent political leader. On December 17, 1967, Australian Prime Minister Harold Holt

From the perspective of a century later, of course, it is obvious that extending federal welfare benefits and unemployment compensation would have helped. But these social programs did not exist at the time, they could not have been put in place quickly, and no one would expect them to be enacted in response to a single local disaster in any case. Thus the idea that the voters blamed Wilson for failing to extend federal disaster assistance, as some readers have suggested to us, is a form of historical presentism—a serious misreading of circumstances at the time.

#### A CENTURY OF DROUGHTS AND FLOODS

Any single instance of blind retrospection, no matter how dramatic, may be dismissed as coincidence or anomaly. Our argument is that voters *consistently* and *systematically* punish incumbents for conditions beyond their control. To assess that broader argument we turn to a comprehensive analysis of electoral responses to droughts and floods by voters in presidential elections throughout the 20th century. We show that voters do indeed punish the incumbent party at the polls for presiding over dry or wet spells.

The data on droughts and floods employed in our analysis consist of monthly readings from 1897 to 2000 of hydrological conditions in each of 344 climate divisions covering the contiguous 48 U.S. states.<sup>19</sup> Each observation summarizes daily data from several weather stations within each climate division. We measure wet and dry spells using the Palmer Hydrological Drought Index (PHDI), an index of long-term moisture supply.<sup>20</sup> A PHDI reading of zero represents an ideal moisture level; negative values represent droughts and positive values represent wet spells.<sup>21</sup>

disappeared while swimming in shark-infested waters at Cheviot Beach near Portsea, Victoria. His body was never found.

<sup>19</sup> The data were generated by the U.S. government and are publicly available from the National Climatic Data Center (NCDC), a unit of the National Oceanic and Atmospheric Administration in the U.S. Department of Commerce. See <http://wfw.ncdc.noaa.gov/oa/climate/onlineprod/drought/readme.html>; also [http://ingrid.ideo.columbia.edu/SOURCES/NOAA/NCDC/CIRS/ClimateDivision/dataset\\_documentation.html](http://ingrid.ideo.columbia.edu/SOURCES/NOAA/NCDC/CIRS/ClimateDivision/dataset_documentation.html).

<sup>20</sup> We believe that the Palmer Hydrological Drought Index provides a better measure of the damage associated with droughts and floods than the Palmer Drought Severity Index, which measures the severity of dry or wet spells of weather rather than long-term moisture supplies.

<sup>21</sup> PHDI values between 2 and 3 represent "moderate" droughts, values between 3 and 4 represent "severe" droughts, and values less than 4 represent "extreme" droughts, and similarly

We aggregate the monthly data for each climate division by computing the absolute value of the sum of monthly PHDI readings from May through October of each calendar year. For simplicity we refer to the result of this calculation as a "drought index," but it is important to bear in mind that the absolute values reflect both wet and dry spells.<sup>22</sup> We further aggregate the data to the level of states by computing a simple average of the annual absolute PHDI values for the climate divisions in each state.<sup>23</sup>

The result of these calculations is an index of climatic pain running from 0.04 to 49.08, with a mean value of 11.03 and a standard deviation of 6.29. Low values of the index are good and high values are bad for voters and thus, according to our account, for incumbent presidents. Our 4,992 observations (for each of 48 states in each of 104 years) include 649 (13%) with absolute PHDI values in excess of 18, the equivalent of a full year of "severe" drought or wetness; 203 observations (4%) have absolute PHDI values in excess of 24, the equivalent of a full year of "extreme" drought or wetness.

We investigate electoral responses to droughts and floods by conducting statistical analyses of popular support for incumbent party candidates in 26 presidential elections (from 1900 through 2000). Thus our analysis is not based on a single, possibly idiosyncratic drought or flood. Rather, we examine an entire century of wet and dry spells, relying upon the random occurrence of numerous droughts and floods to distinguish their common effects from potentially confounding specific circumstances. Our analyses employ state-level voting data and six different versions of our drought index.<sup>24</sup> In each

for positive values indicating wet spells. The distribution of PHDI values is approximately normal, with no asymmetry apparent between the severity of wet spells and dry spells.

<sup>22</sup> This calculation assumes that equally severe droughts and wet spells are equally painful to voters. We investigated that assumption by repeating our statistical analyses using separate measures of droughts and wet spells. The estimated effects were generally similar. For example, in the simplest regression model presented in the first column of table 5.2, distinguishing between droughts and wet spells produced estimated effects of .067 (with a standard error of .047) and .066 (with a standard error of .050), respectively.

<sup>23</sup> Most states are composed of between seven and nine climate divisions; eight states have one, two, or three divisions. The climate division boundaries sometimes reflect geographical features such as coastal areas or mountain ranges, but more often follow county lines.

<sup>24</sup> All our statistical analyses weight each state in each year by the number of votes cast in the presidential election; thus, populous states and those with heavy turnout get more weight than those with fewer voters, and more recent elections get more weight than those earlier in the century covered by our analysis.

Table 5.2. Droughts, Floods, and Presidential Elections, 1900-2000

	Drought index			Rural drought index		
	(1)	(2)	(3)	(4)	(5)	(6)
Election-year drought index	-0.060 (0.031)	-0.052 (0.034)	—	-0.176 (0.083)	-0.140 (0.082)	—
(Election-1) drought index	—	-0.043 (0.029)	—	—	-0.116 (0.088)	—
(Election-2) drought index	—	0.016 (0.036)	—	—	0.023 (0.102)	—
(Election-3) drought index	—	-0.043 (0.040)	—	—	-0.024 (0.102)	—
Time-weighted drought index	—	—	-0.104 (0.045)	—	—	-0.273 (0.122)
Standard error of regression	3.61	3.60	3.60	3.61	3.61	3.60
Adjusted R <sup>2</sup>	.88	.88	.88	.88	.88	.88
N	1,233	1,233	1,233	1,233	1,233	1,233

Parameter estimates (with standard errors in parentheses) from ordinary least squares regression analyses of incumbent vote (%) by state; states weighted by turnout; observations clustered by election year. Election-specific intercepts and election-specific effects of lagged incumbent vote, twice-lagged incumbent vote, % rural and South not shown.

case, we expect droughts and floods to depress the incumbent party's popular vote share. Table 5.2 presents the key results.

In order to allow for other factors affecting the incumbent party's fortunes in each state in each election, we take account of the incumbent party's vote share in the same state in each of the previous two presidential elections, the percentage of the population living in rural areas, and an indicator variable for southern states. The effects of all of these factors are allowed to vary from one election to the next, so that there are 130 coefficients in each regression model in addition to those reported in table 5.2—an intercept for each

election, a coefficient for the lagged incumbent vote in each election, and so on. Only the effects of drought are assumed to be constant across elections.

The simplest version of our analysis, reported in the first column of table 5.2, employs the absolute PHDI value for each state in each election year as the primary explanatory factor. The negative estimated effect indicates that, on average, voters punished incumbent parties for droughts and wet spells; the *t*-statistic for this parameter estimate is -1.9, so the effect of wet and dry spells on election outcomes cannot easily "be dismissed by the dubious as a coincidence" (Barnhart 1925, 529).<sup>25</sup> Nor is the estimated effect trivial in magnitude. It implies that wet or dry conditions in a typical state and year (an average absolute PHDI value of 11) cost the incumbent party 0.7 percentage points, while "extreme" droughts or wet spells (absolute PHDI values of 24 or more) cost incumbents about 1.5 percentage points.

The second column of table 5.2 reports the results of a slightly more complicated analysis in which the drought index values for all four years of each president's term appear as separate explanatory factors. Here, the estimated effect of election year drought is quite similar to the estimated effect in the first column, and drought values in two of the three preceding years appear to have additional (albeit slightly smaller) negative effects. In the third column we employ a time-weighted cumulative drought index in which drought conditions in each year of a president's term get twice as much weight as those in the preceding year.<sup>26</sup> Once again, the effect of droughts and wet spells on the incumbent party's vote share is clearly negative (in this case with a *t*-statistic of -2.3) and of considerable magnitude (costing the incumbent party about 1.1 percentage points in a typical state and year).

The remaining three columns of table 5.2 repeat the analyses reported in the first three columns, but with each drought variable multiplied by the proportion of the population living in rural areas in each state and year. The resulting rural drought indices allow for the possibility that wet and dry spells

<sup>25</sup> The observations in the regression analyses reported in table 5.2 are clustered by election year, which allows the unmeasured factors affecting incumbent party support in each year to be correlated across states. The result of clustering is to increase the estimated standard errors (and reduce the associated *t*-statistics) by about 35%.

<sup>26</sup> The resulting weights attached to drought index values in the four years leading up to each election are .06667, .13333, .26667, and .53333. Deriving geometrically declining weights from the separate estimates reported in the second (or fifth) column of table 5.2 would produce roughly similar weights.



may be particularly consequential in rural areas where farming, ranching, and forestry are major economic activities. However, allowing for the difference in scales between the original and rural drought indices, the pattern of estimated effects turns out to be quite similar. For example, the estimated effect of election year rural drought in the fourth column implies that the incumbent party lost 0.6% of the vote in a typical state and year (as compared with 0.7% in the first column).

The strength and consistency of these results across a variety of analyses employing different versions of our drought index should leave little doubt that droughts and wet spells *in general* had a negative effect on electoral support for the president's party.<sup>27</sup> Climatic distress is a pervasive risk to the reelection chances of every incumbent, and no more controllable than the rain.

An important disadvantage of the summary results presented in table 5.2 is that they conceal a great deal of potentially interesting variation in effects across election years, some of which may be attributable to more or less effective governmental responses and some of which may reflect other factors. For example, as noted by Barnhart (1925, 536–539), insufficient rainfall has less impact on livestock ranchers than on farmers. Thus we expect that some droughts will have substantial economic and political impacts and others less so, depending on where they occurred. That variation is conveyed by figure 5.4, which presents separate estimated effects of election-year drought on the incumbent party's vote share in each election.

The estimated effects of droughts and wet spells are clearly quite variable, with almost half of the election-specific estimates more than twice as large—and a few as much as five times as large—as the corresponding overall estimates in table 5.2. A detailed examination of those varying responses might shed very useful light on the psychology and sociology of voters' attributions of responsibility for natural disasters. However, that sort of detailed examination is beyond the scope of the present study.

Rather than attempting to provide a detailed analysis of climatic retrospection in each election, we propose here merely to emphasize that our analysis of droughts and floods cannot be dismissed as a bit of Dust Bowl

27 In addition to the variety of regression analyses reported in table 5.2 we examined models with separate effects for droughts and wet spells, models with nonlinear variants of our drought indices, models allowing for secular trends in the magnitude of drought effects, models allowing drought effects to vary with prior partisanship, and models employing interactions between local climatic conditions and national climatic conditions. All of these models produced clear evidence of drought effects, but none added significantly—in terms of statistical fit or substantive insight—to the simpler analyses reported in table 5.2.

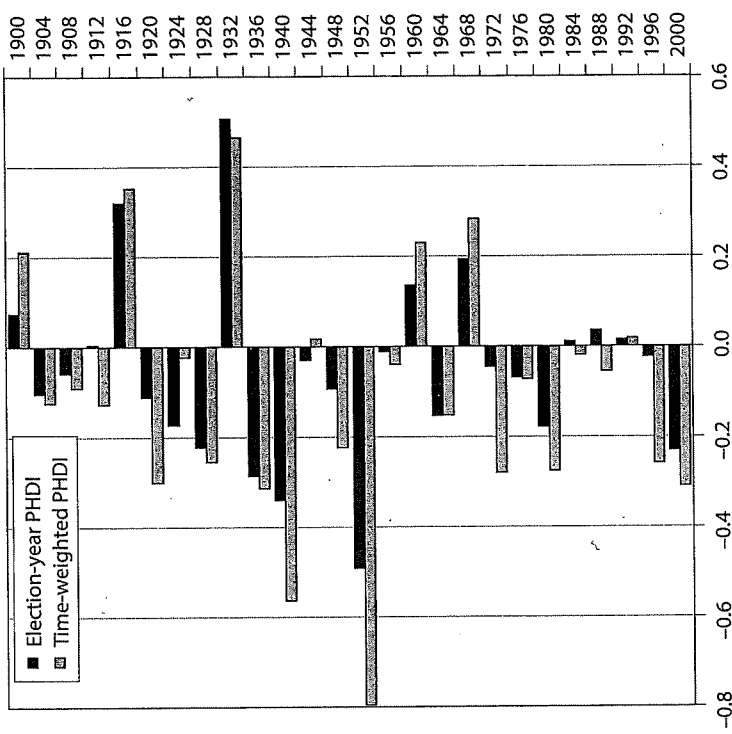


Figure 5.4. Election-Specific Estimates of Drought and Flood Effects, 1900–2000

antiquarianism. We do so by examining the electorate's response to droughts and floods in the 2000 presidential election. The 2000 election occurred under relatively unfavorable climatic conditions for the incumbent party. The average absolute PHDI value was about 10% higher than the historical average, with severe drought in parts of the South and West and excessive wetness in the Dakotas and New York and Vermont.<sup>28</sup>

28 Drought conditions were most severe in Arizona and Alabama, which had drought index values in excess of 20; Louisiana, Montana, Georgia, Mississippi, Texas, Utah, Wyoming, New Mexico, and Nevada also had drought values in excess of 15. At the opposite extreme, the Dakotas, Vermont, and New York had absolute PHDI values ranging from 15.8 to 20.3.

Table 5.3. The Effect of Drought on the 2000 Presidential Vote

	(1)	(2)	(3)
Election-year drought index	-0.231 (0.073)	—	—
Rural drought index	—	-0.546 (0.259)	—
Time-weighted drought index	—	—	-0.310 (0.103)
1996 Clinton vote (%)	0.915 (0.113)	0.896 (0.119)	0.802 (0.115)
1992 Clinton vote (%)	0.206 (0.121)	0.237 (0.127)	0.291 (0.120)
Rural (%)	-0.098 (0.026)	-0.032 (0.040)	-0.116 (0.027)
South	-0.60 (0.76)	-0.96 (0.79)	-1.57 (0.68)
Intercept	-0.14 (3.33)	-3.14 (3.40)	3.40 (3.86)
Standard error of regression	1.94	2.06	1.96
Adjusted R <sup>2</sup>	.92	.91	.92
N	48	48	48

Parameter estimates (with standard errors in parentheses) from ordinary least squares regression analyses of Democratic vote (%) by state; states weighted by turnout.

Table 5.3 presents our analysis of the impact of climatic conditions on the 2000 presidential vote. We consider three different versions of our drought index, with statistical controls for previous Democratic votes, percentage rural, and an indicator variable for southern states. All three versions of the statistical analysis fit the data very well, with average errors of less than two percentage points in the predicted Democratic vote share and adjusted R<sup>2</sup> values in excess of .90.

The results of these analyses suggest a great deal of continuity with previous partisan voting patterns, albeit with some significant defections from the Democratic ticket in rural states. The three versions of our drought index all have strong negative estimated effects on the Democratic vote, with *t*-statistics ranging from -2.1 to -3.2. The magnitudes of these estimated effects suggest that the Democrats' national vote share was from 1.6 to 3.6 percentage points lower than it would have been had voters not been inclined to make the incumbent party feel their climatic pain. The aggregate effect implied by the best-fitting model, employing the simple election year drought index, falls in the middle of that range at 2.7 percentage points.

This estimate implies that 2.8 million people voted against Al Gore in 2000 because their states were too dry or too wet. As it turned out, Gore could have used those votes. Attributing them to the various states in proportion to their drought scores suggests that climatic retribution cost Gore seven states—Arizona, Louisiana, Nevada, Florida, New Hampshire, Tennessee, and Missouri—and almost three times as many electoral votes as Florida's infamous "butterfly ballot" (Wand et al. 2001). Voters responded to climatic distress in 2000, as they have repeatedly throughout the past century, by punishing the incumbent government at the polls.

#### HOLDING INCUMBENTS RESPONSIBLE: WHY SO MUCH PUNISHMENT?

When voters endure natural disasters they generally vote against the party in power, even if the government could not possibly have prevented the problem. In our experience, this simple fact induces in many readers a powerful urge to defend the sophistication and rationality of the electorate. Here we take up their arguments.

First, it is possible that voters did not blame the government for the disasters themselves, but did blame it for *exacerbating* or failing to *ameliorate* the damage. In that case, retrospection might not be blind. The point is not a new one. As Barnhart (1925, 540) wrote of 19th-century Nebraska,

To suggest that the farmer held the politician responsible for the shortage of rainfall would be an unwarranted exaggeration of the thoughtlessness of the voters. But it is quite another matter to suggest that the drought in Nebraska made a bad set of agricultural conditions worse and that the politicians were held responsible for some of the conditions. Perhaps some held them responsible for most of them. The situation of

many farmers forced them to think about the things that had brought about that situation. . . . They could not make it rain, but they thought they could lower railroad rates.

The difficulty with this argument is its strong policy component. If the voters learn in disasters what they had hitherto not suspected—namely that stronger government intervention in the economy is needed—then droughts ought to push electorates to the ideological left. What they actually do, however, is reduce support for incumbents regardless of their ideological commitments. “Throwing the rascals out” after droughts did lead to left-wing gains in Nebraska in the 1890s; but where left-wing governments have been in power the lack of rain has created surges of right-wing voting, as in the American Dust Bowl in the mid-1930s. Similarly in the case of the sharks, if New Jersey voters in 1916 wanted a better government social safety net enacted, then their swing from Wilson to his Republican opponent was politically quite illogical.

Moreover, whatever the voters learn in natural disasters has a very short half-life. As many scholars have noted, Populism declined rapidly as economic conditions improved in the farming states. Confirming that historical pattern, table 5.2 suggests that droughts in the first half of a president's term have no consistent influence on the voters. Similarly, by 1920 the Jersey Shore was back to its customary partisanship among New Jersey counties, and the beach areas of Ocean County that had suffered most from the shark attacks were 8 percentage points more favorable to the party of Wilson than the near-beach, just the reverse of their views at the time of the attacks four years earlier. In sum, the voters quickly forget their grievances. Short-term anti-incumbent voting without substantial policy content is the only common pattern in the electorate's response to natural disasters.

An alternative defense of voter rationality is that the electorate punishes incumbents not for the *occurrence* of natural disasters, which are clearly beyond their control, but for insufficient *responses* to those disasters. In this view, voters monitor incumbents' performances in the aftermath of disasters in much the same way, and for much the same reasons, that they monitor other aspects of incumbent performance—imperfectly, based on observed results, in order to *select* competent leaders and provide *incentives* for future performance. This sort of punishment seems wholly consistent with the logic of retrospection we set out in chapter 4.

The problem with this interpretation is that it is manifestly irrational for well-informed voters to punish incumbent politicians for droughts, floods,

or other natural disasters *in general*. Any *particular* drought, flood, or other natural disaster may be an occasion for rational punishment of incumbent politicians who fail to take adequate steps to ameliorate its effects. However, a competent electorate must recognize that incumbents' preparations for, and reactions to, the substantial physical and social dislocations resulting from major disasters are, by definition, better than average about half the time. And while it may be sensible for voters to punish incumbents who perform less well than average in the wake of a disaster, punishing those who perform *better* than average is counterproductive both from the standpoint of selecting competent leaders and from the standpoint of providing proper incentives for future incumbents to do their best under difficult circumstances.

Thus, we should expect to find rational electorates *rewarding* incumbents for better-than-average responses to natural disasters as often as they *punish* incumbents for worse-than-average responses. The evidence strongly contradicts this expectation. Energetic politicians can sometimes successfully claim credit for federal disaster assistance or shift the blame to other levels of government (Abney and Hill 1966; Arceneaux and Strein 2006; Healy and Malhotra 2010; Gasper and Reeves 2011; Brader n.d.). However, the one-sided tendency for voters to punish rather than reward their leaders in times of distress is evident in our statistical results. The estimated drought effects in table 5.2 represent average effects on incumbent vote shares of all the droughts and floods of the 20th century, including effective responses as well as ineffective ones. If reward and punishment were distributed impartially, the *average* effects of droughts and floods would be zero. Instead, they are preponderantly negative. Voters are much more apt to punish their leaders than to reward them.

In human terms, the voters' behavior is understandable. They see friends and neighbors pitching in to help immediately after the disaster. They do not understand why the government cannot do the same. To citizens, government bureaucracies with their rules to prevent fraud and their accounting regulations may dispense disaster assistance with a heartless attitude and a lethargic spirit. Or, if the money is dispersed more freely, citizens may complain about waste and abuse (Schneider 1995, 70–71). Either way, in the wake of a disaster the government will look inept or uncaring to a devastated and emotionally shaken electorate. Hence, the voters will punish most of the time. But doing so can neither relieve their distress nor produce more competent or caring political leaders.

One final version of the principal-agent argument is more consistent with our evidence. On this view, voters simply punish incumbent leaders any time

their own well-being falls below "normal" levels, regardless of whether the incumbents have performed well or badly. Disasters are very likely to cause hardship even when incumbents' responses are vigorous and competent, so this sort of decision rule leads to frequent punishing after disasters, thus matching the evidence. And it is "rational" in the technical sense set forth in chapter 4—if voters are incapable of distinguishing relevant from irrelevant sources of subjective well-being, the best they can do may be to respond identically to both. In that case, voters will respond in much the same way to shark attacks and droughts as they do to poor tax policies and disastrous foreign wars. Incapable of assessing causation, they will mechanically translate pain into electoral punishment. But that is just what we mean by blind retrospection.

#### THE SOCIAL CONSTRUCTION OF BLAME

Thus far we have written of retrospective voting as if hardship itself created electoral backlash. Like many other scholars, we have de-emphasized the voters' interpretation of their plight, as if it had little causal importance. In fact, however, we believe that voters' attributions of blame are often crucial in their decisions to punish incumbents. Positive or negative events that voters themselves would recognize as politically irrelevant, such as the outcomes of local college football games, may also influence voting behavior through subconscious effects on voters' moods (Healy, Malhotra, and Mo 2010). But substantial punishment at the polls is likely to be grounded in a belief, however farfetched, that the government is somehow responsible for the voters' pain.

It is easy to overlook the need for social interpretation of hard times, since suggestions about their meaning are typically mass-produced. Political and ideological entrepreneurs have an incentive to construct explanations and solutions, often self-serving, for people's hardships. Amplified by the mass media, these ideas may increase or decrease the likelihood that citizens will attribute responsibility for social problems to the government (Iyengar 1991). In garden-variety economic recessions, the accepted stories about blame are familiar, and the process of generating common understandings occurs so smoothly and easily that its importance may go unnoticed. Natural disasters, by contrast, create deeper and unfamiliar hardships, which lead to uncertainty and even fear. The old complacent assumptions are shown to be mistaken, and a search begins for new explanations that will avoid a repetition of the disaster (Cantril 1958, chap. 1; Birkland 1997). People are ready to listen.

After disasters, the more popular attributions of blame and proposals for reform often come from widely trusted sources and appeal in a clear, simple way to broadly shared values, though not necessarily those that intellectual elites rely on for their political judgments. As Hadley Cantril (1941, 67) put it, "There are short-cut rationalizations which fire the imagination and spread because they somehow express the dissatisfactions from which people have been suffering and at the same time imply a new direction and purpose." If a single nutty or dangerous vision comes to be sufficiently widely shared, demagogues may be able to ride it to power.

Elite culture is usually (though not always) less susceptible to nutty or dangerous visions. It may even play some role in discouraging the most ignorant or vicious attributions of blame. But popular culture is never entirely under the control of the respectable. A variety of unconventional interpretations and nostrums may be available, and under the right circumstances deviant doctrines may attract considerable popular acceptance. Some medieval towns blamed the plague on Jews, prostitutes, beggars, or foreign agents (Herlihy 1997, 65–67). Some New Jersey residents in 1916 thought that German U-boats might have induced the sharks to attack (Fornicola 2001, 166–170). Some Americans in the grip of the Spanish Influenza pandemic two years later feared that "plague germs were inserted into aspirin made by the German drug company Bayer" (Kolar 1999, 3).

Different sectors of the population, immersed in distinct subcultures, may find different explanations appealing. Ideological commitments may color the plausibility of alternative explanations, as with the Federalists' and Republicans' competing accounts of the yellow fever epidemic of 1793. Physicians "divided bitterly over the cause of the epidemic," with Republicans generally attributing it to poor sanitation, climatic conditions, and the unhealthy location of Philadelphia, while Federalists blamed disembarking refugees from Haiti; in fact, "both sides were right" (Pernick 1972, 562–563). If available interpretations are sufficiently contested, and if incumbents can exploit competing explanations to exonerate themselves and blame others, they may sometimes escape blame altogether (McGraw 1991; Arceneaux and Stein 2006).

In other cases, blame may fail not because there are too many available interpretations of disaster but because there are too few. In 1874, for example, locust swarms devastated large swaths of western Nebraska and adjacent states. By fall, many farmers literally faced starvation. The Army had clothing and food supplies stored in the area, but refused to distribute the clothes

until several weeks after the fall election, and did not give out food until the following February (Lockwood 2004, 80–84). Nonetheless, the incumbent Republicans sailed to victory in Nebraska in 1874, and repeated plagues of locusts throughout the mid-1870s did not notably dent their popularity in either gubernatorial or presidential elections (Nebraska Legislative Reference Bureau 1918, 436–506). The voters did not punish. The simplest explanation is probably that in the thinly populated farming areas of Nebraska at that time, communication was poor and no shared interpretation of the disaster emerged. A strong ethic of self-reliance also militated against expecting assistance from the government (Lockwood 2004, 38–39). And, perhaps most important, the Populists were not well organized until a decade later and did not mount a serious campaign for governor until 1890. At that point, farmers suddenly had a credible explanation for their troubles and a target for their frustrations—and punishment began.

If our interpretation of the cultural element in natural disasters is correct, then it should be possible to point to a major disaster for which a government was plausibly responsible, yet for which it escaped electoral blame because the case for responsibility was never constructed by political opponents. The Spanish Influenza pandemic of 1918 represents a remarkable case of just that sort. The magnitude of the disaster was epic; most estimates of the worldwide death toll range from 20 million to 40 million, with some as high as 100 million. In the United States alone, the flu killed approximately half a million people—more than the total number of American battle deaths in World War I, World War II, Korea, and Vietnam combined (Crosby 1989; Kolata 1999, 285, ix-x). This was no mere blown college football game. If voters punished the incumbent government whenever they felt significant unhappiness, the millions of people who lost friends or family members to influenza in 1918 would have produced the greatest anti-incumbent landslide in American electoral history. But electoral retribution requires voters to imagine, however plausibly or implausibly, that incumbent leaders could have prevented or ameliorated their pain. In the case of the flu pandemic, that crucial attribution of political responsibility was lacking. As a result, as best we can tell, the electorate utterly failed to respond to the greatest public health catastrophe in U.S. history.

The 1918 midterm election occurred just as the pandemic was at its peak in many parts of the country, with flu deaths numbering more than one thousand per week in some major cities. Using detailed data on influenza mortality rates from 16 states and 29 major cities (Crosby 1989; Pyle 1986, 46–47), we examined voting patterns in the 1918 midterm election, looking for evidence

of electoral retribution aimed either at Democratic gubernatorial candidates or at incumbent governors regardless of party.<sup>29</sup> We also examined the 1920 presidential vote. In no case did we find reliable statistical evidence that voters in the worst-hit states and cities punished anyone at the polls.<sup>30</sup>

One important race was almost certainly affected by the pandemic—the Senate contest in New Mexico, in which President Wilson clumsily attacked Republican Albert B. Fall at the same time Fall was grieving over the deaths from influenza of two of his children. Fall was elected by fewer than 2,000 votes, and Alfred Crosby (1989, 175) quite plausibly argued that “sympathy for the bereaved Fall caused Wilson’s attack to backfire.” In this isolated case, the horrific effect of the pandemic became a potent political issue; but in the country as a whole, remarkably, it did not.

It is impossible to know, even in retrospect, how much could have been done to minimize the loss of life in what was, after all, a vast and virtually unprecedented tidal wave of human misery. Nevertheless, it seems clear that a rational electorate could reasonably have held its leaders accountable, in part, for the devastating consequences of this natural disaster. Even with due allowance for the less developed public health technology of 1918, there is little reason to doubt that tens of thousands of flu victims could have been saved by more effective government action. Efforts to stem the contagion, or even to track its spread, were slow and disorganized (Crosby 1989, chaps. 1, 2, esp. 49–51; Kolata 1999, 10, 19, 22–23).

So why no electoral retribution? For one thing, blaming the government was not easy: the country was at war, making criticism seem unpatriotic. Both the government and the press downplayed flu risks (Barry 2004, chap. 29). Indeed, the pandemic seems to have received remarkably little national attention. As one historian put it, “When you talk to people who lived through it,

29 As it happens, 13 of the 16 “registration states” for which detailed mortality data are available had Republican governors in 1918. Thus, voting patterns would look very different depending on whether voters chose to punish incumbent governors or the party of the Democratic president.

30 In most cases, we examined the impact of influenza mortality rates in the final four months of 1918; for some cities we also had more detailed data that allowed us to examine the impact of flu deaths in the weeks immediately preceding the election. Not surprisingly, given the limitations of the available data, all of our statistical results were fairly imprecise and, in some cases, quite sensitive to changes in the sample or variable definitions. On the whole, however, it seems clear that the flu pandemic had little or no political effect. Some of the estimates suggest, quite implausibly, that incumbent governors actually *gained* votes in the major cities with the highest death rates. Only the estimate for Democratic votes at the state level had the “correct” (negative) sign, and even that estimate was of very modest magnitude.



they think it was just their block or just their neighborhood" (Crosby, quoted by Kolata 1999, 8). Victims were widely scattered around the country; and since people died of influenza every year, no one could be certain that their own spouse or parent or child was one of the "excess deaths" from the epidemic, much less a death that the government might have prevented.

Most important, no thread of elite rhetoric or popular discourse seems to have suggested any attribution of responsibility to President Wilson or other public officials. As long as no one supplied a convincing argument that the government did control or should have controlled the spread of the pandemic or its horrific consequences, the pain of millions failed to have any electoral impact. President Wilson was berated for the insufficiency of his efforts to stem the tide of shark attacks in New Jersey in 1916 and taunted with editorial cartoons featuring shark fins; but there is no evidence of a comparable outcry over his handling of the flu pandemic, except in an isolated instance in which he insensitively attacked a political opponent whose children had been among the victims.

The striking absence of a broad-based electoral response to the flu pandemic dramatically illustrates the importance of voters' cultural understandings of causation and responsibility. In the language of Deborah Stone's (1989, 283) typology of causal frameworks, voters thought of the pandemic as part of the natural world ("the realm of fate and accident") rather than as part of the social world ("the realm of control and intent"). Obviously, such cultural understandings are subject to change.<sup>31</sup> But at the time, while hundreds of thousands of people died, no one thought to blame the pharaoh.

## CONCLUSION

In most recent scholarly accounts, retrospective voting is a natural and rational feature of democratic politics. In our view it is natural, but not so obviously rational. Indeed, blind retrospection of the sort we have documented in this chapter seems to us to provide a significant challenge to the conventional understanding of political accountability in modern democracies.

We have shown that voters sometimes punish incumbent political leaders for misfortunes that are clearly beyond the leaders' control. Moreover, we have

31. Modern governments have certainly believed that the political cost of a major flu epidemic might be considerable, as witnessed by the Ford administration's aggressive—as it turned out, overtly aggressive—response to the swine flu scare in the 1970s (Neustadt and Fineberg 1983).

shown that they do so with considerable regularity. The fact that American voters throughout the 20th century punished incumbent presidents at the polls for droughts and floods seems to us to rule out the possibility that they were reacting to subpar *handling* of misfortunes rather than to the misfortunes themselves. After all, it is hard to see how incumbent presidents' handling of droughts and floods could have been substantially worse than average over the course of an entire century.

Of course, voters may themselves contribute to poor disaster preparedness by insisting on low taxes and less intrusive government. In that case, government performance in response to disasters will nearly always seem poor in some absolute sense, and incumbents will be punished accordingly. But this sort of retrospective punishment is self-defeating in exactly the way we have suggested, since the randomness of the punishment from the standpoint of incumbents makes it pointless (in an electoral sense) for any incumbent to invest *ex ante* in adequate preparations for disasters (Healy and Malhotra 2009).

What, if anything, is wrong with blind retrospection? In a world of great uncertainty and costly attentiveness, perhaps this is exactly what voters should do to hold their leaders accountable—"only calculate the changes in their own welfare," as Fiorina (1981, 5) put it, and vote accordingly. Maybe there was something Woodrow Wilson could have done for the Jersey Shore, even if no informed person at the time could think of what that might be. And if a few pharaohs perish needlessly as a result of fanciful causal chains in the voters' minds, is that such a high price to pay for a system in which every incumbent has a strong incentive to do whatever she can to maximize her citizens' welfare? In short, aren't voters behaving rationally when they reward or punish incumbents for good or bad times?

In one sense, this view of retrospective voting is quite right. When voters are utterly ignorant about whether and how their leaders' actions affect their own welfare, blind retrospection may be "rational" in a narrow, technical sense. However, that does not imply that it will be sensible or prudent. Crazy beliefs can make crazy behavior "rational." But as the models presented in chapter 4 demonstrate, ignorance about reality can be quite costly in the realm of democratic politics, just as it is in other aspects of life.

Our account strikes directly at the heart of the common normative justification for the retrospective theory of political accountability. In that view, while voters may know very little, they can at least recognize good or bad government performance when they see it. Thus, they can retrospectively reward or punish leaders in a sensible way. We agree that voters operating on the basis of a valid, detailed understanding of cause and effect in the realm of public

policy could reward good performance while ridding themselves of leaders who are malevolent or incompetent. But real voters often have only a vague understanding of the connections (if any) between incumbent politicians' actions and their own well-being. Even professional observers of politics often struggle to understand the consequences of government policies. Politics and policy are complex. As a result, retrospective voting is likely to produce consistently misguided patterns of electoral reward and punishment.

To sensibly translate an assessment of economic or social *conditions* into an assessment of political *performance*, citizens must find—and accept—a valid cultural understanding of the causal relationships linking the actions of public officials to changes in the public's welfare. When is one such understanding accepted rather than another? A healthy democratic culture among political elites can, no doubt, help significantly to constrain the scapegoating impulses of democratic electorates. Yet just as much or more seems to depend on the political folk culture among ordinary citizens, or on different folk subcultures for different groups.

Tracing how a specific political attribution of blame attains plausibility among inattentive citizens suddenly in want of an explanation for their troubles is a daunting task. The young and old, the rich and poor, the educated and uneducated are all swept along by the ideas popular in their groups, and sometimes all are swept along together. Certain looks, certain sounds, and certain arguments meet widespread needs in a particular culture at a particular time, nearly always for complex reasons unforeseeable in advance. The only certainty is that there is nothing very rational about the process.

Our analysis suggests that "blind" retrospection on the basis of overall well-being, with no consideration of the impact of government policies on that well-being, is very unlikely to provide much in the way of effective accountability, notwithstanding the fact that it may be "rational" in a narrow sense. Voters ignorant about evidence and causation, but supplied with a tale of incumbent responsibility, will punish incumbents whenever their subjective well-being falls below some fixed standard, regardless of whether or not their pain is in fact traceable to the incumbents' policies.

The "rough justice" (Fiorina 1981, 4) embodied in the electoral verdicts rendered by such voters is likely to be very rough indeed. And the rougher it is, the less incentive reelection-minded incumbents will have to exert themselves on the voters' behalf. As a result, voters who cannot distinguish the effects of shark attacks and droughts from the effects of tax policies and foreign wars are likely to experience more than their share of misguided tax policies and disastrous foreign wars. This sort of voting is hard to square with rosy in-

terpretations of retrospective accountability, and even harder to square with the folk theory of democracy, in which ordinary citizens assess their public life critically, weigh the qualifications of competing candidates for public office, and then choose between the candidates in accordance with their own values.

Democracies take their electoral direction from human beings with limited capacities for self-government. Human passions remain powerful, and human understanding remains weak. Under sufficient pressure, voters sometimes lash out blindly. Such events are not quaint historical footnotes rendered irrelevant by modern education and hygiene. Indeed, in just the past century many citizens—and many prominent intellectuals—have been enthusiastic supporters of Nazis, Bolsheviks, Mao's Communist guerrillas, and a host of other brutal demagogues whose policies seemed to offer attractive solutions to fundamental social problems that the previous incumbent rulers had failed to master.

Blind retrospection afflicts us all. It is the inevitable consequence of bewildering social complexity and human cognitive limitations—limitations that the rise of democratic government has not altered. The conventional account of retrospective voting, minimalist as it is, fundamentally underestimates the limitations of democratic citizens and, as a result, the limitations of democratic accountability.

are assumed to respond to income growth only in the last two quarters before the election. Certainly none provides empirical support for the notion that voters weigh economic outcomes equally, or even approximately equally, over the entire period for which the incumbent president and his party could plausibly be considered responsible.

#### CAN MYOPIC VOTERS SELECT COMPETENT ECONOMIC MANAGERS?

Our analysis of economic voting suggests that presidential elections provide significant moments of economic accountability. As Gerald Kramer (1971, 140) put it decades ago, "election outcomes are in substantial part responsive to objective changes occurring under the incumbent party." However, objective changes in economic well-being seem to matter significantly only if they occur in close proximity to Election Day. In this respect, economic accountability is seriously skewed by voters' short time horizons. Myopic voters reward their elected leaders for *some* good times and punish them for *some* bad times. Is that good enough?

We explore this question by assessing separately the two distinct rationales for retrospective voting set out in chapter 4—selection and sanctioning. In this section we consider the selection rationale, in which retrospective voters systematically improve the average economic competence of political leaders by retaining incumbents who are competent economic managers and replacing those who are not. In the next section we turn to the sanctioning rationale, in which retrospective voters induce leaders (regardless of their intrinsic competence) to strive to produce good economic outcomes by rewarding success and punishing failure.

The selection rationale requires us to assume that there are real differences in economic competence between competing political teams. If there are, it certainly seems reasonable to suppose that an administration's past economic performance might provide voters with a useful clue about its competence, and thus about its likely future economic performance. However, in light of the crucial importance of this assumption for the whole notion of retrospective selection, it is striking that it has never, as far as we know, been subjected to any systematic empirical examination. Proponents of retrospective voting have simply *assumed* that there are real, persistent differences in economic competence between competing teams of political elites, and that voters who retain or replace the incumbent administration on the basis of its past performance are likely to get more competent government in the future as a result.

In this section, we attempt to put those assumptions to the test. In particular, we explore how well an electorate voting on the basis of very recent economic performance and ignoring or discounting most of what has happened on the incumbent's watch—an electorate of the sort portrayed in our empirical analysis of economic voting in modern presidential elections—is likely to succeed in selecting competent economic managers. The answer depends not only on the behavior of voters, but also on the dynamics of real income growth.

Three important characteristics of income growth in the post-war United States are evident in figure 6.2, which shows *annualized* quarter-by-quarter changes in real disposable income per capita over the entire period covered by our analysis. First, it is hard to see any striking trends in prevailing growth rates over this period.<sup>16</sup> Second, there is a great deal of short-term volatility in growth; quarterly fluctuations of 5 percentage points or more in the annualized growth rates are fairly common.<sup>17</sup> Third, and related, there is surprisingly little temporal continuity in quarter-to-quarter growth rates, even over short periods of time. For example, relating income growth in the current quarter to income growth in the four previous quarters produces lag coefficients of .00, .06, -.00, and -.14, with an adjusted  $R^2$  statistic of .01. Thus, current deviations from the long-run average growth rate are very unlikely to provide a good basis for forecasting future deviations—a point of crucial but little-noticed importance if retrospective voting is supposed to result in the selection of successful future leaders.<sup>18</sup>

The solid line in figure 6.2 represents the average income growth rate for each four-year presidential administration.<sup>19</sup> Here, as in the analyses reported in table 6.1, we treat each administration's economic responsibility as beginning five months after it takes office (with Q3) and ending five months after the next inauguration (at the end of Q18).

<sup>16</sup> More elaborate statistical analysis reveals some drift over time in the average income growth rate, but the discernible trends are of modest magnitude.

<sup>17</sup> To some extent these fluctuations presumably reflect measurement error. However, there is no strong pattern of negative association between growth rates in successive quarters, as one would expect if the *level* of income in each quarter were measured with substantial random error.

<sup>18</sup> Simply regressing the current rate of real income growth on the previous quarter's rate produces a lag coefficient of -.03 and an adjusted  $R^2$  statistic of -.00. More elaborate time series analyses produce similar results.

<sup>19</sup> We ignore the midterm transitions from Kennedy to Johnson and from Nixon to Ford, since in both cases there seems to have been a good deal of continuity in economic management teams and policies.

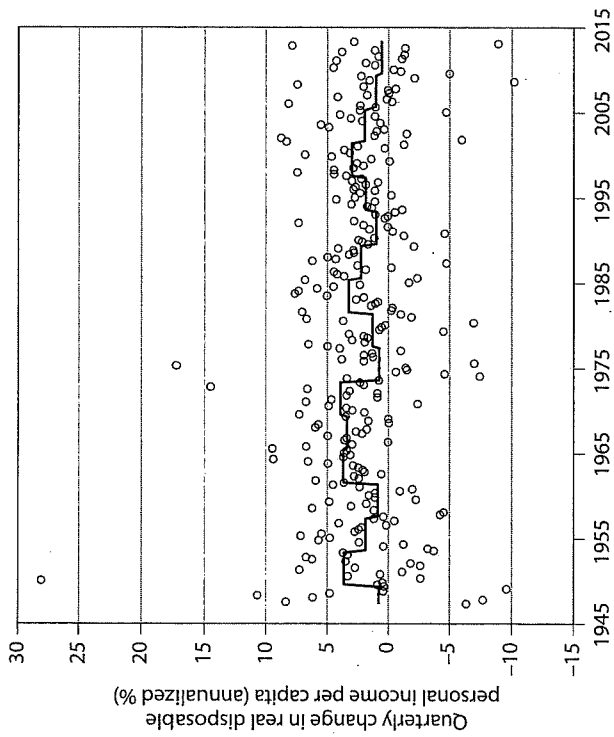


Figure 6.2. Income Growth by Presidential Administration, 1947–2013

Changes in the average level of real income growth from one administration to the next are clearly discernible in figure 6.2; but they are modest by comparison with the short-term fluctuations from one quarter to the next.<sup>20</sup> As a result, the correlation between each quarter's change in income and the average change for the corresponding administration is only .29. Allowing for the mechanical contribution of each quarter's change to the corresponding average reduces that correlation to only .16. (The correlation between quarterly and administration-average changes in GDP is only slightly more impressive.) More elaborate analyses including multiple lagged values, polynomial trends, and seasonal adjustments produce essentially similar results.

These features of the data are relevant for our analysis of retrospective voting because they suggest that voters are likely to have a good deal of difficulty discerning systematic changes in growth rates from one administration to the

<sup>20</sup> It is not even possible to decisively reject the null hypothesis that the average income growth rates for every administration are identical; the tail probability for an  $F$ -test is .12.

next amid the constant short-term fluctuations in their economic well-being. Obviously, if voters cannot reliably discern differences in economic outcomes they cannot reliably respond to those differences.

The volatility of short-term income growth rates underscores the difficulty of making sensible assessments of an incumbent's economic competence on the basis of how things are going right now. Voters evaluating the economy on the basis of any one or two quarters are likely to do little better than chance at capturing the potentially meaningful differences in long-term performance represented by the administration average growth rates depicted in figure 6.2. In particular, the correlation between income growth over any two quarters and income growth over the same president's entire term is only .41, implying that more than 80% of the variance in overall economic performance is unrelated to short-term successes or failures of the sort that actually matter in the voting booth. Of course, part of the relevant overall performance in our accounting occurs *after* Election Day—but that simply underscores the extent to which any inference about the differential competence of specific administrations to produce future economic growth can be little more than a roll of the dice.

The fundamental dynamics of income growth suggest that myopic voters are unlikely to succeed in detecting genuine differences in the economic competence of specific administrations. However, suggesting that effective retrospective selection is likely to be difficult is not the same thing as establishing that it does not occur. Unfortunately, it is notoriously hard to provide a direct test of the prospective efficacy of economic voting, since the logic of the argument hinges on an inherently unobservable comparison between future economic performance under the incumbent and future economic performance under the challenger. Inevitably, the voters' choice on Election Day relegates one of those two outcomes to the realm of historical might-have-beens. Thus, we are reduced to testing the plausibility of the retrospective selection hypothesis from a variety of more or less oblique angles.

We begin with a test inspired directly by the model of retrospection set out in chapter 4. In that model, the voter's *electorally relevant subjective well-being* is represented as the sum of incumbent competence and random factors. In keeping with our evidence regarding myopic economic voting, we shall assume here that electorally relevant subjective well-being is indexed by observed income growth over a six-month period. In that case we can use observed economic data to directly estimate the magnitude of random factors contributing to subjective well-being. Table 6.3 presents the results of two different versions of such an analysis.

Table 6.3. Discerning the Economic "Competence" of Presidential Administrations Based on Six-Month Growth Records

	Presidencies	Parties
Competence (deviations from average income growth)	+0.26, -0.24, +1.38, +0.51, -0.89, +0.30, -1.26, +0.20, -1.16, -0.34	+0.41 (Dem), -0.28 (Rep)
Standard error of regression	2.98	2.98
Adjusted $R^2$	-.000	.005
$N$	131 (distinct six-month windows)	
Average efficiency of selection (improvement over random chance) (%)	8.3	9.1

Ordinary least squares regression parameter estimates for six-month income growth rates regressed on indicators for parties or presidencies, 1947–2012.

In the first version of the analysis, we assume that competence attaches to specific incumbents; in that case, our best guess regarding the competence of each president is represented by the average rate of income growth over that president's time in office. In the second version of the analysis, we assume that competence attaches to political parties rather than individual incumbents, and is constant over the entire post-war era; in that case, our best guess regarding the competence of each party is represented by the post-war average rate of income growth under all presidents of that party. In both cases, we assume—with unrealistic optimism—that competence is perfectly measured by the observed average rate of income growth, ignoring other aspects of competence as well as effects of good or bad luck that extend over an entire presidency or over all the Democratic or Republican presidencies of the post-war era.<sup>21</sup>

21 Treating observed average growth rates as *estimates* of underlying competence would clearly make retrospective voting look even less effective, since it would add another layer of uncertainty to voters' inferences based on short-term economic performance.

In keeping with these assumptions, we relate income growth in each six-month period of the post-war era—successive slices of economic experience of the sort actually relied on by myopic economic voters—to indicator variables representing the ten distinct presidencies (in the first version of the analysis) or Democratic and Republican parties (in the second version of the analysis). The results of these analyses are presented in the left- and right-hand columns of table 6.3, respectively.

In the analysis assuming incumbent-specific competence, the average (annualized) income growth rate range from 1.4 percentage points higher than the overall average (for Kennedy-Johnson) to 1.3 percentage points below the overall average (for George H. W. Bush). If we assume that voters assessing each incumbent contemplated replacing him with a challenger of average quality, these deviations from overall average growth rates directly reflect each incumbent's presumed relative competence.<sup>22</sup> However, the statistical analysis implies that six-month slices of short-term income growth are virtually uncorrelated with (these estimates of) underlying competence. The adjusted  $R^2$  statistic, a measure of the signal-to-noise ratio, is slightly less than zero, while the average error of the regression is 3.0, implying a "noise" level (literally) off the chart of retrospective difficulty depicted in figure 4.1.<sup>23</sup>

In the analysis for parties, average income growth under Democratic presidents was .4 percentage points higher than the overall average, while average growth under Republican presidents was .3 percentage points lower than the overall average. Thus, our best estimate of the economic competence of Democratic presidents (relative to Republicans) is +.7, while our best estimate of the economic competence of Republicans (relative to Democrats) is -.7. Even these estimates capitalize on chance to some degree, though less than in the case of individual administrations. Nevertheless, the statistical analysis accounts for only a tiny fraction of the substantial variation in short-term income growth rates. The adjusted  $R^2$  statistic is .005, implying that the best estimate of the signal-to-noise ratio for this model is essentially zero.

What do these results suggest about the likely effectiveness of retrospective selection? Clearly, the electorate's likely success in inferring the competence

22 A more complicated analysis might allow for long-term trends in income growth rates. Doing so would make the voter's problem even harder, since incumbent-specific deviations in growth rates would be (slightly) smaller in magnitude, making incumbents look (slightly) less distinct.

23 Given our scaling of variables, the standard error of the regression implies an estimated value of slightly less than 3.0 for the "noise" parameter  $\sigma$  in figure 4.1, well off the 0–2 scale.



of parties or specific incumbents from narrow slices of economic experience is essentially nil. Translating the estimates of relative competence and the estimated magnitudes of random factors into probabilities of reelection (as in figure 4.1), the average improvement over random chance is in each case less than 10%. That is, voters are less than 60% likely to elect a party or candidate who is actually more competent than the available alternative, and more than 40% likely to elect a party or candidate who is *less* competent than the available alternative.

These meager estimated success rates are partly a reflection of the fundamental difficulty of the voters' task—and it is worth reiterating that we have simplified that task considerably for purposes of our analysis by assuming that good and bad luck always cancels out over the course of a presidency, leaving average observed income growth as a perfectly reliable indicator of competence. However, the estimated inefficiency of selection is also attributable in significant part to myopia, which impels voters to judge incumbents on the basis of brief, relatively uninformative slices of economic experience.

Another way to examine the effectiveness of retrospective voting is to see what happens *after* each election. If we take seriously the notion that reelection hinges on economic competence, one implication is that we should expect to see more economic growth when the incumbent party is reelected than when it is dismissed by the voters. In the former case the incumbent party has presumably been retained because its past performance makes it a better than average bet to provide good economic management in the future.<sup>24</sup> In the latter case the new administration is presumably an unknown quantity, a random draw from some underlying distribution of economic competence. A secondary implication of this logic is that future economic performance should be less variable when the incumbent party is retained, since reelected administrations are a truncated subsample of the underlying distribution of economic competence (the worst economic performers having presumably been weeded out at reelection time).

These implications of the retrospective selection model are examined in table 6.4, which compares post-election economic performance in cases where the incumbent party has been reelected with the corresponding performance in cases where the incumbent party has been voted out of office. Our measures of post-election economic performance are the average rates of growth in real disposable income per capita and GDP per capita in Q19–Q34, the

<sup>24</sup> Obviously, election outcomes are influenced by many other factors besides economic management. That fact complicates our proposed comparison, but does not alter the basic logic.

Table 6.4. Post-Election Economic Growth under Old and New Administrations

	Incumbent party retained	Incumbent party replaced	Difference
POST-ELECTION INCOME GROWTH			
Average growth	7.95 (1.73)	9.17 (1.70)	-1.22 (2.43)
Standard deviation	4.89	4.82	+0.06
POST-ELECTION GDP GROWTH			
Average growth	7.87 (2.50)	8.46 (1.43)	-0.59 (2.88)
Standard deviation	7.06	4.06	+3.00
<i>N</i>	8 (1948, 1956, 1964, 1972, 1984, 1988, 1996, 2004)	8 (1952, 1960, 1968, 1976, 1980, 1992, 2000, 2008)	16

Average values (with standard errors in parentheses) and standard deviations of post-election (Q19–Q34) growth in GDP per capita and real disposable income per capita (%), 1948–2008.

entire four years of the new president's economic stewardship. (We lose 2012 from this comparison because Barack Obama's full second-term post-election economic performance is not yet known; but we add 1948, giving us eight cases of reelection and eight cases of replacement.)

Expectations derived from the retrospective selection perspective fare quite poorly in this comparison. Post-election income growth was slightly more variable under reelected incumbents than in cases of partisan turnover, and post-election GDP growth was much more variable—the opposite of what one might expect if voters succeeded in weeding out poor performers. More important, the average rates of post-election income and GDP growth were *lower* in terms with reelected incumbents than in cases of partisan turnover—by 13% and 7%, respectively. Of course, with only 16 administrations to analyze, these comparisons are far from definitive.<sup>25</sup> However, there is

<sup>25</sup> The estimated difference in post-election income growth between reelected administrations and those that were replaced is -1.22 percentage points (with a standard error of

clearly no support here for the notion that incumbents are retained or replaced on the basis of intrinsic economic competence.

A further, more stringent test of the retrospective selection perspective is to examine directly whether the factors that affect post-election economic performance, whatever they may be, also affect voters' decisions to retain or replace the incumbent party. If voters succeeded to some extent in forecasting future economic competence, and vote accordingly—as the logic of the retrospective selection perspective implies—then we should observe more electoral support, other things being equal, for incumbents who turn out to be competent economic managers after the election than for those who turn out to be incompetent.

Of course, the context of post-election policy-making may bear little resemblance to what voters imagine on Election Day. The prospect of changes in the global economy, wars and terrorist attacks, unforeseen bouts of inflation or recession, and alterations in the makeup of the incumbent management team all produce a great deal of uncertainty about future economic performance under any incumbent. But that uncertainty is by no means irrelevant to our assessment of the likely efficacy of retrospective selection. To argue that the test of actual post-election performance is too stringent is, in effect, to concede that retrospective selection is too difficult for voters to succeed at it.

Table 6.5 presents the results of additional statistical analyses intended to search for traces of economic foresight in the outcomes of recent presidential elections.<sup>26</sup> Our aim in each case is to account for the reelected incumbent party's popular vote margin (in percentage points) on the basis of *post*-election income or GDP growth, measured over the entire four-year period for which the reelected administration is assumed to be responsible (Q19–Q34).

The statistical analysis reported in the first column of table 6.5 tests whether voters' support for incumbent administrations depends at all upon the *future* success of those administrations in producing income growth. The answer seems to be no: the estimated effect of future income growth is actually negative, though the estimate is quite imprecise. Of course, this statistical

2.42). The estimated difference in post-election GDP growth is  $-0.59$  percentage points (with a standard error of 2.88). Going back to 1929 (using *annual* income data to approximate the quarterly data) adds four additional observations to the analysis, but leaves the results essentially unchanged.

26 By the logic of retrospective selection, new administrations are simply random draws from some overall distribution of competence—mere campaigning sheds no light on their specific competence. Thus, we limit our analyses here to the eight elections in which the incumbent party was, in fact, reelected—the instances in which pre-election economic performance should, theoretically, have provided useful insights regarding post-election performance.

Table 6.5. Economic Voting on the Basis of Post-Election Income and GDP Growth

	Post-election income growth		Post-election GDP growth	
	(1)	(2)	(3)	(4)
Post-election growth (Q19–Q34)	-0.11 (0.68)	-0.51 (0.59)	0.17 (0.46)	0.11 (0.56)
Short-term income growth (Q14–Q15)	—	6.73 (2.69)	—	4.92 (3.45)
Incumbent tenure (years)	—	-1.72 (0.61)	—	-1.80 (0.67)
Intercept	13.66 (6.19)	13.39 (4.60)	11.48 (4.76)	12.61 (5.00)
Standard error of regression	8.73	5.70	8.65	6.17
Adjusted $R^2$	-16	50	-14	42
$N$	8 (1948, 1956, 1964, 1972, 1984, 1988, 1996, 2004)			

Ordinary least squares regression parameter estimates (with standard errors in parentheses) for reelected incumbent party's popular vote margin (%), 1948–2008.

analysis would be very hard to defend as a plausible model of voting behavior, and its poor fit to the data (with an adjusted  $R^2$  statistic of  $-0.16$ ) does nothing to inspire confidence.

The analysis presented in the second column of table 6.5 adds two explanatory factors that we have already shown to be strongly related to incumbent parties' electoral fortunes—short-term income growth in the two quarters preceding the election and the length of time (in years) that the incumbent party has controlled the White House. Not surprisingly, adding these factors to the analysis improves the goodness of fit substantially. The average error is reduced by more than one-third, and the adjusted  $R^2$  statistic increases from  $-0.16$  to  $.50$ .<sup>27</sup> However, the key parameter estimate, for post-election income

27 The estimated effects of short-term income growth and incumbent tenure in this truncated sample are reassuringly similar in magnitude to the estimates for the full sample

growth, remains perversely negative and even increases in magnitude. In short, our analysis provides no support for the notion that retrospective voters can reliably recognize and reward competent economic management.

The third and fourth columns of table 6.5 present the results of parallel analyses focusing on post-election GDP growth. They require little by way of exposition, since the results are generally similar to those for post-election income growth. In the third column of the table, without controls, it appears that voters are no more supportive of incumbents who turn out, upon reelection, to preside over high rates of GDP growth. Adding control variables does nothing to improve the estimated effect of future GDP growth on incumbent vote margins. As with the results presented in the first two columns of table 6.5—and those presented in table 6.4—these results cast doubt on the notion that incumbents are reelected, even in part, on the basis of economic competence.

It behooves us to reiterate that the tests of retrospective selection offered here are indirect, and that their statistical power is quite limited. Thus, our results certainly do not *prove* that successful retrospective selection through economic voting is impossible. Nevertheless, it is striking how little empirical trace we have been able to find of the presumed connection between the choices of American voters in presidential elections and their subsequent economic fortunes. We eagerly await better data and more powerful tests. In the meantime, however, the efficacy of retrospective economic voting as a mechanism for selecting competent leaders must, we think, be taken as a matter of theoretical speculation, not as a matter of fact.

#### CAN MYOPIC VOTING PROVIDE INCENTIVES FOR SUSTAINED ECONOMIC GROWTH?

As we saw in chapter 4, even if retrospective voters are unlikely to succeed in selecting competent economic managers, they may contribute to their post-election welfare by incentivizing incumbents—regardless of competence—to do whatever they can to produce economic growth. The key idea here is that a rational electorate may have good reason to punish the incumbent party when times are bad, despite the fact that the past is past, and regardless of *why* times are bad, simply in order to discipline *future* incumbents. Whether or not voters can reliably identify variation in the competence of incumbents, their

of post-war elections reported in table 6.1, though of course they are much less precise due to the reductions in both sample size and variation in the dependent variable.

tendency to vote on the basis of perceived success or failure may reinforce incentives for future incumbents to do whatever they can to maximize voters' welfare.

In models of this sort, voters have no reason to expect the outcome of any particular election to have any direct impact on the course of the economy; whichever party wins will produce exactly the same post-election economic conditions. From this perspective, the empirical results presented in tables 6.3 and 6.4 are no embarrassment to an electorate whose sole aim is to establish a reputation for rewarding success and punishing failure. Retrospective voting, in this interpretation, is a simple mechanism for enforcing as much correspondence as possible between the goals of the voters and the goals *any* set of self-interested reelection-seeking politicians will actually pursue once they are in office.

For this retrospective sanctioning model to work effectively, however, voters must presumably reward (or punish) current incumbents for doing (or failing to do) what they want future incumbents to do. The usual, and quite sensible, assumption is that voters want to maximize the present discounted value of their future income stream. If we add the corollary assumption—well supported by a good deal of empirical evidence—that gains in income tend to cumulate and persist over time,<sup>28</sup> then the increment in voters' long-term economic welfare on any given incumbent's watch can be indexed, to a good approximation, by the average rate of real income growth throughout his term.<sup>29</sup> Thus, that is what rational voters implementing a retrospective sanctioning strategy should reward or punish.

It is striking, then, that the substantial empirical literature on economic voting finds voters doing no such thing. Most analysts have simply focused, without any explicit explanation or justification, on economic conditions in the year of the election. And, as we have shown here, paying more careful attention to earlier conditions clearly demonstrates that they are, to a good

28 In time series terms, income is not mean-reverting. Either a unit root or trend stationarity would imply that condition, with parameters varying by presidential administration. The voter's inference problem is somewhat different in the two cases, but the distinction makes no difference for our purposes.

29 A more exact calculation depends on the discount rate voters use to assess the present value of future income. The average rate of real income growth throughout a president's term is insensitive to differences in the timing of that income growth, which affect voter's welfare in the short run but are swamped in the long run, assuming that there is a long run and that voters do not discount future income too severely. To the extent that voters are impatient, they should reward early income growth more than later income growth of similar magnitude. As we have seen, they do not. In any case, for the sake of simplicity we adopt the customary assumption in the literature—that voters care about long-run average income growth.

economy flourish around election time—"a bias toward policies with immediate, highly visible benefits and deferred, hidden costs—myopic policies for myopic voters" (Tufté 1978, 143). However, a good deal of subsequent empirical research has produced less clear-cut results. As one observer (Schultz 1995, 79) put it, "while the general logic behind the theory is quite persuasive, the empirical evidence for electoral-economic cycles is spotty at best."

We believe that a variety of problems contribute to the spotty empirical support for electoral cycles in economic outcomes. The power of statistical tests to detect pre-election fluctuations in economic conditions may be quite modest in short time series with few elections. Some studies focus on secondary measures of economic performance such as unemployment and inflation rather than income growth, which is clearly of primary electoral importance. Others are insensitive to important institutional details, such as whether the timing of elections is fixed or variable. The advent of "rational expectations" theory in macroeconomics has probably contributed to the skepticism of the scholars generating the empirical evidence, despite demonstrations that political business cycles of some form can emerge even in dynamic models with fully rational voters and politicians (Rogoff and Sibert 1988; Rogoff 1990).

We do not pretend to offer a thorough review and assessment of empirical evidence regarding political business cycles here. Instead, we provide a very simple test of whether economic performance has fluctuated in the run-up to recent presidential elections in the way one might expect if incumbent presidents were catering to myopic economic voters. The results of this test are reported in figure 6.3, which summarizes average growth rates in real GDP per capita and real disposable income per capita in each year of the presidential election cycle, from inauguration (year 1) to reelection (year 4), from 1949 through 2012. The shaded bars represent GDP growth, while the solid bars represent income growth.

We have two distinct reasons to expect a stronger electoral cycle in real disposable income growth than in GDP growth. First, since voters are much more responsive to income growth than to GDP growth, incumbents have much more to gain from manipulating income growth. Second, as a practical matter, it is probably a good deal easier for incumbents to manipulate income growth—for example, through transfer payments and tax cuts—than to manipulate the pace of real economic output through macroeconomic policy.

As one might expect, given these considerations, the most striking aspect of figure 6.3 is the marked increase in average income growth (though not in GDP growth) in presidential election years. The changes in income growth over the course of the election cycle are impressively large, with average

approximation, electorally irrelevant—despite being (at least) as important as election-year conditions for voters' long-term economic well-being.

How would we expect reelection-seeking incumbent politicians to respond to the electoral incentives generated by such "myopic" retrospection? The obvious-seeming answer is that they should attempt to maximize income growth in the immediate run-up to elections, but care little about what happens to the economy at other times. A president who shirks (or, more realistically, pursues his own ideological agenda) in the months just before the election may be punished, but a president who shirks (or pursues his own ideological agenda) earlier in his term is likely to suffer little or no penalty at the polls. Thus, there is little or no *electoral* incentive for presidents to promote myopic voters' well-being during much of their time in office. Meanwhile, voters' short time horizons magnify incentives for incumbents to manipulate the economy in order to maximize economic performance around election time. The result is "a rational incentive for the party in power to manipulate the business cycle for electoral benefit" (Erikson 1989, 570).

William Nordhaus (1975) first called scholars' attention to the fact that voters' short time horizons might induce a "political business cycle" in democratic political systems, with incumbents' efforts to stimulate the economy producing regular peaks in growth around election times. The first extensive empirical analysis of political business cycles, by Edward Tufté (1978), provided statistical evidence of electoral cycles in transfer payments, income growth, unemployment, and inflation, as well as considerable qualitative evidence of specific efforts by incumbents to produce those cycles. Richard Nixon in 1972 was a particularly energetic manipulator of everything from the money supply (through his erstwhile political ally Arthur Burns, the chairman of the Federal Reserve Board) to effective dates of increases in Social Security benefits and payroll taxes; as Tufté (1978, 63) delicately put it, "The extremes of 1972 were special because Richard Nixon was special."<sup>30</sup>

Tufté worried that political manipulation of economic policy could generate significant social costs due to wasteful government spending and other forms of "economic instability and inefficiency" aimed at making the

30 According to Tufté, Nixon's enthusiasm for political manipulation of the economy arose in significant part from his unhappy experience with an *un*manipulated economy in his first presidential campaign in 1960. In a post-campaign memoir, Nixon (quoted by Tufté 1978, 6) wrote, "In October, usually a month of rising employment, the jobless rolls increased by 452,000. All the speeches, television broadcasts, and precinct work in the world could not counteract that one hard fact: Nixon lost the official national popular vote by fewer than 120,000 votes. (See Gaines 2001 for a recount.)"

success or failure. Across the 16 administrations for which both numbers are available, the former exceeds the latter by an average of about 0.2 percentage points.<sup>32</sup> Given the estimated impact of short-term income growth on incumbent vote margins in table 6.1, this difference suggests that post-war incumbent parties have routinely padded their electoral margins by about two-thirds of a percentage point through political manipulation of the economy.

Of course, in any particular instance the difference between short-term and long-term economic performance is likely to reflect sheer good or bad luck in the timing of economic growth. We note, however, that Richard Nixon's 1972 reelection bid—Tufte's (1978) prime source of colorful examples of economic manipulation—produced one of the largest differences on record between short-term income growth and long-term GDP growth, boosting Nixon's vote margin by 7 or 8 percentage points. (Ronald Reagan in 1984 and Lyndon Johnson in 1964 probably gained 8 or 9 percentage points.) At the opposite extreme, Jimmy Carter's vote margin in 1980 was probably *reduced* by 10 or 11 percentage points—more than enough to cost him reelection—by the fact that voters judged his economic record on the basis of an election-year recession rather than on the basis of his overall economic performance.

More generally, our comparison of politically consequential short-term income growth rates with long-term GDP growth rates suggests that economic booms in the run-up to elections are not wholly a matter of good or bad luck. In the ten elections since World War II in which incumbent presidents ran for reelection, the annual rate of income growth in the run-up to the election exceeded the annual rate of GDP growth over the president's entire term by an average of almost half a percentage point, producing an estimated increase in the incumbents' average vote margin of 1.5 percentage points. However, in the six elections in which the incumbent president was retiring—giving him a less direct stake in the election outcome—the election year's income growth rate was slightly *lower* than the long-term GDP growth rate, reducing the successor candidates' average vote margin by about 0.7 percentage points.<sup>33</sup>

32 This difference is primarily due to presidential elections rather than to the distinction between GDP and income growth rates. Outside of presidential election years, the average rate of GDP growth (1.94) slightly exceeds the average rate of income growth (1.77); but during presidential election years the average rate of income growth (3.03) greatly exceeds the average rate of GDP growth (2.04).

33 Omitting the 1968 election from this calculation (on the grounds that incumbent president Lyndon Johnson was eligible to run for reelection but chose not to) reduces but does

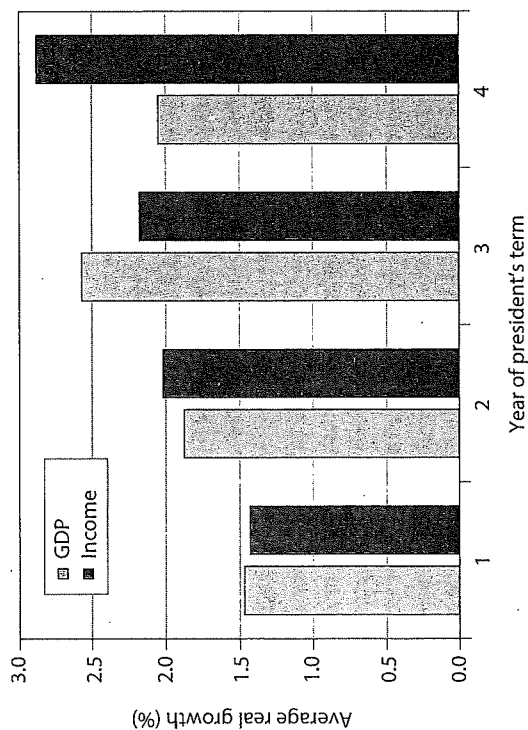


Figure 6.3. GDP and Income Growth through the Electoral Cycle, 1949–2012

income growth a full percentage point—more than 50%—higher in presidential election years than in non-election years (and substantially *lower* in the year *after* each election).<sup>31</sup> Moreover, the average election-year income growth rate exceeds the corresponding GDP growth rate by 40% (2.9 versus 2.0 percentage points)—just as we would expect if incumbents tailored their economic policies to appeal to an electorate that is much more attentive to election-year *income* growth than to election-year *GDP* growth. These pronounced electoral cycles of economic growth are a testament to the incentives created by a myopic electorate. As Tufte (1978, 137) insisted almost 40 years ago, “economic life vibrates with the rhythms of politics.”

Another way to gauge the magnitude of election-related economic manipulation is to compare the average (annualized) rate of income growth in Q14 and Q15 with the average rate of GDP growth in Q3 through Q18—arguably the best single measure of each administration's overall economic

31 The *t*-statistic for the difference in average income growth between election years (2.9) and non-election years (1.9) is 2.8, making the difference exceedingly unlikely to be due to chance.



The social cost of this political manipulation of the economy is very difficult to assess. If incumbents merely shift income gains that would otherwise have occurred earlier or later into the election-year window of voter cognitive, the result is simply to add one more electoral edge to every incumbent's balance sheet.<sup>34</sup> On the other hand, to the extent that political manipulation of the economy generates deadweight losses due to wasteful government spending and other forms of "economic instability and inefficiency" (Tufte 1978, 143), the costs are presumably borne by the same voters whose shortsightedness generates the electoral economic cycle in the first place.

Less institutionalized, more volatile national economies may shed clearer light on the consequences of election-time pork. A director of the Turkish Treasury noted that the inflation rate was 17% in the six months preceding the November 1987 elections, and 48% in the next six months as voters paid the price for pre-election pump priming. The corresponding rates were 26% and 36% at the October 1991 elections and 24% and 46% at the December 1995 elections. "We, economy bureaucrats, do not like early and frequent elections," he said (Uras 2004).

What, if anything, can be done to mitigate political manipulation of the economy? Nordhaus (1975, 188–189) doubted "the practical possibility" of what he referred to as "the 'classical' political solution" — "improve the information available to voters so they can judge and condemn the partisan nature of myopic economic policies." He argued instead for tinkering with the frequency of elections or assigning greater control over economic policy to "persons who will not be tempted by the Sirens of partisan politics," including unelected central bankers and interest groups. In contrast, Tufte (1978, 154) concluded his analysis by asserting that "sleazier efforts at manipulating economic policy for short-run advantage cannot survive public scrutiny," and hoping that books like his might "improve the level of public understanding so that voters can evaluate and repudiate corrupt economic policies."

not eliminate this apparent penalty. Conversely, the apparent election-year boost for incumbents seeking reelection would be almost twice as large (increasing their average vote margin by 2.8 percentage points) if we excluded Jimmy Carter's election-year recession in 1980.

<sup>34</sup> Of course, sensitive democrats might still consider it unfair or disasteful that the most shameless manipulators get the biggest electoral edge. And sensitive Democrats might still consider it unfair or disasteful that Republicans have generally benefited from uncharacteristically high short-term income growth when they have held the White House (gaining an average of 2 percentage points) and from uncharacteristically low short-term income growth when Democrats have held the White House (gaining an average of 1 percentage point). Barrels (2008, chap. 4) provided a more detailed analysis of the partisan implications of myopic economic voting in U.S. presidential elections.

Alas, there is little evidence to suggest that developments in the decades since Tufte wrote have done anything to increase public scrutiny of electoral economic cycles. For example, in the period from 1949 through 1976, the average income growth rate was 1.2 percentage points higher in presidential election years than in other years; since 1977 it has been 1.1 percentage points higher in presidential years than in other years.<sup>35</sup> Voters continue to think and act in much the same way they always have—and politicians behave accordingly.

## CONCLUSION

We have argued for an understanding of retrospective voting quite different from the conventional scholarly view that we summarized in chapter 4. While we agree that voters' assessments of economic conditions play a substantial role in determining election outcomes, we part company with those who see retrospective voting as a mechanism for ensuring effective political accountability.

The evidence presented in this chapter casts considerable doubt on the view that citizens can reliably form and act upon sensible retrospective judgments at election time. While they vote on the basis of how they feel at the moment, they forget or ignore how they have felt over the course of the incumbents' term in office. Like medical patients recalling colonoscopies, who forget all but the last few minutes (Kahneman 2000; Redelmeier, Katz, and Kahneman 2003), the voters' assessments of past pain and pleasure are significantly biased by "duration neglect." Their myopia makes retrospective judgments idiosyncratic and often arbitrary.

The myopic behavior of economic voters does not seem to be a mere reflection of complex campaign environments or restless media coverage. Similar biases are evident even in simple laboratory settings where experimental subjects have (modest) material incentives to make sensible choices. For example, one such study found "evidence of three important deviations from optimal retrospective": subjects "overweighted recent performance," "allowed unrelated events that affected their welfare to influence evaluations of incumbents," and "were influenced by rhetoric" to focus on recent rather than cumulative performance (Huber, Hill, and Lenz 2012, 738). The authors concluded that their findings point to "inherent limits in citizens' ability to motivate incumbent performance" (Huber, Hill, and Lenz 2012, 739).

<sup>35</sup> The *t*-statistics for the earlier and recent differences are identical, 1.9.

The result of this kind of voter behavior is that election outcomes are, in an important sense, *random*. Again, as we said in chapter 1, we do not mean to suggest that outcomes are random in the literal sense of being utterly chaotic or unpredictable. Indeed, as figure 6.1 highlighted, they are highly structured. Our point is that the most important single factor in determining who wins—myopic retrospection—is, from the standpoint of democratic accountability, essentially arbitrary. It is not much related to either ideology or incumbent performance. Thus, when electoral competition is sufficiently vigorous for the outcome to be in doubt, the choice between competing governing teams is likely to hinge on the accident of whether “the economy is flourishing in the final weeks of a campaign” (Uchitelle 2004). Usually, in effect, the voters toss a coin.

One last possibility remains, however, by which retrospective voting might be rehabilitated, at least in part. Perhaps in times of national crisis, voters can rise to the occasion, casting an ideologically sophisticated vote that changes the political system dramatically, ushering in a new majority party with a mandate for dramatic policy change (Sundquist 1983). The period from 1932 to 1936 is the strongest recent nominee for a watershed of that kind. The next chapter therefore takes up the elections of the Franklin Roosevelt years.

## CHAPTER SEVEN

## A Chicken in Every Pot: Ideology and Retrospection in the Great Depression

For their part, academics have tended to dismiss campaign slogans of the past like “the full dinner pail” and “a chicken in every pot” on the grounds that something deeper must have been going on in these elections. But perhaps it wasn’t.

—David R. Mayhew, *Electoral Realignments* (2002, 161)

Americans are accustomed to thinking of the New Deal realignment as a triumph of both democratic responsiveness and Democratic Party ideology. The Great Depression following the stock market crash of 1929 had brought unprecedented economic catastrophe on the nation; real income per capita fell by an appalling 28% from 1929 to 1932. A rigidly conservative Republican government resisted public pressure to provide energetic relief and institutional reforms. Republicans lost heavily in the 1930 congressional elections. Then in the presidential election of 1932, voters responded with a historic repudiation of the incumbent president, Herbert Hoover. Franklin Roosevelt swept into office with 57% of the popular vote, and the Democrats—a minority party for most of the preceding 70 years—won 313 of the 435 seats in the House of Representatives.

Roosevelt’s first hundred days in the White House brought a flurry of innovative policies. A robust economic recovery followed in short order. Real incomes increased by one-third over the course of Roosevelt’s first term. Unemployment declined by one-third. Voters rewarded Roosevelt with a landslide reelection in 1936; he won more than 60% of the popular vote and carried 46 of 48 states. He went on to win an unprecedented four terms in