Dominic Lusinchi

"President" Landon and the 1936 *Literary Digest* Poll

Were Automobile and Telephone Owners to Blame?

The disastrous prediction of an Alf Landon victory in the 1936 presidential election by the Literary Digest *poll is a landmark event in the history of American survey research in general and polling in particular. It marks both the demise of the straw poll, of which the* Digest *was the most conspicuous and well-regarded example, and the rise to prominence of the self-proclaimed "scientific" poll. Why did the* Digest *poll fail so miserably? One view has come to prevail over the years: because the* Digest *selected its sample primarily from telephone books and car registration lists and since these contained, at the time, mostly well-to-do folks who would vote Republican, it is no wonder the magazine mistakenly predicted a Republican win. This "conventional explanation" has found its way into countless publications (scholarly and in the press) and college courses. It has been used to illustrate the disastrous effects of a poorly designed poll. But is it correct? Empirical evidence, in the form of a 1937 Gallup poll, shows that this "conventional explanation" is wrong, because voters with telephones and cars backed Franklin D. Roosevelt and because it was those who failed to participate in the poll (overwhelmingly supporters of Roosevelt) who were mainly responsible for the faulty prediction.*

Polling and survey research more generally are data collection technologies widely used today in the social sciences. Many apprentices of these disciplines will take a college course in research methods and be told what constitutes a good survey and how to recognize a bad one. Most likely, as an illus-

tration of the latter category, the instructor will recount the story of the 1936 *Literary Digest* presidential poll: it predicted the wrong winner (Alf Landon, the Republican challenger), and, to top it all off, its forecast error was huge (nearly 20 points). They will also be told most probably that the reason for this fiasco was the magazine's reliance on telephone directories and car registration lists to select its sample,[1] thereby biasing it, in the midst of the Depression, in favor of the better-off portion of the electorate, which was bound to vote for the Republican nominee.

Budding social scientists are not the only ones who will have heard this story. Just like baseball, polling is now entrenched in the American culture; the general public is inundated with polls, especially during an election year. From time to time, this tale will appear in the press (e.g., Higgins 2010) and in other publications accessible to the lay public (e.g., Zunz 1998: 64).

But is this account correct? And how did it come about? The answer to the first question will be the primary focus of this article. As for the second question, the explanation for what went wrong with the *Digest* poll was largely crafted by the new players in this arena: the so-called scientific poll takers (Archibald M. Crossley, George H. Gallup, and Elmo Roper). Indeed, the 1936 presidential election witnessed not only the demise of traditional straw polls, chiefly as a result of the *Digest* poll's massive blunder, but the emergence of "scientific" polling. The advent and development of these new polls was made possible by their use of novel sampling methods and by their relative success in predicting the outcome of the election but more specifically by outperforming the *Digest* poll.

Gallup, founder in 1935 of the American Institute of Public Opinion (AIPO), and his fellow pollsters (Crossley and Roper) had a major stake in this story. After their "triumph" over the *Digest* (Rusciano 2007: 315), they were particularly keen to explain the superiority of their approach and show why the old "methods" used by straw polls, of which the *Digest* was the most prestigious example, were fundamentally unsound. Crossley (1937: 27) put it most bluntly: "The *Literary Digest* method is outmoded." In contrast, the new "scientific" polling was capable, they believed, of rendering a true image of America. According to Roper (1940: 326), "Our purpose is to set up an America in microcosm." The superior value of this methodology, as Gallup (1972a: 146) saw it, rested on what he called "scientific sampling procedures"[2] that could yield a *representative* sample of the population of interest:[3] specifically, it could produce a miniature version of the electorate.[4] To pro-

mote the new venture, Gallup had to demonstrate that straw polls in general, and the *Digest* in particular, were based on samples that were biased, that is, *unrepresentative* of the American electorate.[5] All through his career Gallup reiterated his version of the failure of the 1936 *Digest* poll. His authority in the field of polling and survey research was so great that his interpretation has been repeated countless times to college students and others and has pervaded textbooks and manuals on survey research methods and other scholarly works as well as newspaper and magazine stories. Even in the face of empirical evidence contradicting it (Squire 1988), this view persists.

Building on previous investigations (Bryson 1976; Cahalan 1989; Cahalan and Meier 1939; Squire 1988), the focus of this article is to show that the available empirical evidence regarding the causes of this event does not support what has been called the "conventional explanation" (Erikson and Tedin 1981: 953). It will become clear that the *Digest*'s original list was no impediment to predict the correct winner. The reason for the magazine's mistaken prognosis was that those who participated in the poll (the respondents) turned out heavily in favor of the Republican challenger, whereas most of the incumbent president's supporters failed to return their straw ballots—this is referred to as *nonresponse bias*.[6]

This article proceeds as follows. First, I describe how the *conventional explanation* emerged in the aftermath of the poll's blunder and how pervasive this explanation is in the scholarly literature and in other publications that refer to this event. Then, I examine previous attempts to use the only available empirical evidence about the *Digest* poll's failure, and I reanalyze this evidence with techniques advocated by Adam J. Berinsky (2006) when survey results originate from quota-controlled polls of the 1930s. I also look at the response rate of various sociodemographic groups. Next, I discuss the shortcomings of the Gallup poll of 1937 and see if they negate the conclusions of this reanalysis. Finally, I offer some general remarks about the *Digest* poll and propose lessons that can be drawn from it by social science researchers.

The Genesis of the Conventional Explanation and Its Diffusion

A few days before the 1936 presidential election, in which the incumbent Roosevelt was being challenged by Landon, the *Literary Digest*, a weekly magazine with a surprisingly good track record in predicting the outcomes

of previous electoral contests,[7] published the results of its presidential poll: it predicted that Landon would defeat Roosevelt—the former would receive 54 percent of the total popular vote, while the incumbent president would get only 41 percent. In fact, Roosevelt obtained 61 percent of the vote (US Bureau of the Census 1970: 354). Although this severe blunder destroyed the credibility of the magazine, which went out of business two years later, it became a major milestone in the history of polling and survey research in the United States.

What happened? The magazine mailed out in the neighborhood of 10 million "ballots" (postcards) based on a number of sources: the rosters of clubs and associations, voter registration rolls, occupational records, classified mail-order and city directories (*Literary Digest* 1936a: 3–4). But the bulk of its list was based on telephone directories and registers of automobile owners.[8] Of the 10 million cards sent out, nearly 2.4 million were returned, for a response rate of roughly 24 percent. How is it possible that with a sample of that size the magazine blundered so badly?

Much has been written about this event over the years, and many are those who have referred to it. For the purpose of this study, these commentators can be grouped into three categories. First, there are the *protagonists*: those who lived through the event and had a direct stake in explaining what had gone wrong. These include the "scientific" pollsters (Crossley, Gallup, and Roper) who promoted the explanation that was to define the event for years to come. A second group, the *outsiders*, is formed by those who are contemporaries of the protagonists but are located at the periphery of the polling world: they are more disinterested observers. The last group is referred to as the *heirs*, because, for the most part, they simply repeat the explanation of the event advanced by the protagonists and their fellow travelers.

The first to seek an explanation for what had happened to the poll was, not surprisingly, the *Literary Digest* itself. In the November 14, 1936, issue of the magazine, the editors pondered over what had happened: why, having used the "same method" that had been so successful in the past, could they have been so far off the mark?[9] They dismissed emphatically the charge that the *Digest*, because it relied on "telephone books and lists of automobile owners," "simply did not reach the lower strata." Their mailing list was much more comprehensive, and they insisted that the ballots they sent out "*did* reach these so-called 'have-nots' strata."[10] They cited three cities (Chicago; Scranton, Pennsylvania; and Allentown, Pennsylvania) in which they

polled from one-third to all registered voters and obtained the same results as they did at the national level.

The editors had a lot of questions about their poll's failure but could not come up with an explanation. However, some of the questions they asked contained the germ of an answer. For example, they wondered: "Why did only one in five voters in Chicago to whom *The Digest* sent ballots take the trouble to reply? And why was there a preponderance of Republicans in the one-fifth that did reply?" These two questions could have been applied just as easily to the results at the national level, but the editors did not make the connection. Indeed, fewer than a quarter of the ballots mailed out had been returned, and a predominance of them came from Republicans—perhaps this was an indication that the 24 percent who responded to the poll were very different in candidate preference from the rest of the original sample. This hypothesis did not occur to the editors—after all, they were not polling experts in the current sense of the term; besides, polling and survey research was a nascent field. They viewed their poll as a "public service" that gave voice to those who cared to participate. As they declared candidly: "We did not attempt to interpret the figures." They were content with their "scrupulous bookkeeping." They clung to the belief that "in that great mass of post-card votes, representing the opinions of every section, class, age and occupation will be found the answer" (*Literary Digest* 1936b: 5).

If the *Digest* editors could not provide answers, the "scientific" pollsters (Crossley, Gallup, and Roper) and their fellow travelers were ready and willing to do so. Claude E. Robinson (1937: 52–53), who in 1938 became the associate director of AIPO and had studied straw polls extensively (see also Robinson 1932), wrote shortly after the *Digest* fiasco:

> The *principal flaw* in the [*Digest*] poll was that the sample was drawn primarily from higher income groups. In collecting its ballots the *Digest* has always used telephone and automobile names for its mailing list. In a few of the larger cities ballots have been sent to registered voters, but by far the largest percentage of the mailings has gone to owners of telephones and automobiles. (emphasis added)

Fellow travelers in academe, such as Hadley Cantril, would provide further legitimacy to this story. In 1937 he and his colleague Daniel Katz stated: "The *Digest* sent its ballots to names taken largely from readily accessible sources such as telephone books and automobile registration lists. . . . From

the very start, therefore, the poll was loaded toward the upper income level" (Katz and Cantril 1937: 158).

The story was also echoed at the periphery by nonprotagonists but scholarly observers, the *outsiders*. "The cause of the *Literary Digest*'s humiliating defeat was evident," the statistician Frederick F. Stephan (1939: 343) stated emphatically. "Its postcard questionnaires reached and were returned by the predominantly Republican upper and middle income classes in greater proportions than they reached and were returned by the predominantly lower income classes."[11] Fellow statistician Samuel S. Wilks (1940: 266), future president of the American Statistical Association, concurred: "The names of people to be polled were selected from such lists as telephone directories, automobile registration lists, etc. It is quite obvious that such a system of selection will be biased by having too large a proportion of the higher income and social groups, which in the 1936 presidential election were known to be heavily weighted in favor of Landon." "It is well known," Wilks added, "that individuals in the lower income levels are not as responsive to mail ballots as those higher up the scale, which would only accentuate the first kind of bias."

Of the three polling pioneers, Gallup wrote the most and the longest about this event. Throughout his career he mentioned the *Digest* poll in books, articles, and interviews.[12] "With the exception of a few cities," he wrote, "the *Digest* sampled only owners of automobiles and telephones. It neglected for the most part the very poor and those with little purchasing power. . . . The income bias in the *Digest* sample resulted in a disproportionate number of Landon votes" (Gallup 1938: 139). Although other factors in the poll's demise were cited (Gallup and Rae 1940: 48–49), they were considered only minor issues. In fact, as time went by any details regarding the account of the *Digest*'s failure were swept aside (Blondiaux 1998: 259). In the end, what was left of the explanation was reduced to this: "The failure of the *Literary Digest*'s polling approach can be explained rather simply. The *Digest*'s sample of voters was drawn from lists of automobile and telephone owners" (Gallup 1972a: 147).

Given the consensus that quickly emerged on this issue, what reason would one have to doubt what Gallup and other protagonists had to say about the failure of the 1936 *Digest* poll? His position as "the dean of public opinion experts" (Bryson 1976: 184) gave him an aura of infallibility. So it is not surprising that, when a reference was made to the *Digest* poll's debacle, this explanation permeated the social science discourse very quickly.

As early as 1952, one could read in a prominent scholarly journal: "The *chief weakness* of that [the *Digest*] poll was that it did not reach a representative cross section of the electorate. Particularly in the depression years, many people in the lower income groups did not have telephones, and were therefore not listed in the telephone directories. For this reason the ballots were not mailed to a representative cross section" (Smith 1952: 150; emphasis added). This assessment could be found some years later, for example, in the very authoritative *Handbook of Survey Research*: "The *Literary Digest* straw polls of telephone subscribers were based on returns of millions of ballots mailed to all telephone subscribers in the United States" (Rossi et al. 1983: 3). It spread as well to the field of electoral studies and voting behavior. Thus one author states that the *Digest* sample was made up of the "attentive middle class" drawn "from telephone books and automobile registration lists" (Shively 1972: 621). Also, one authority on US elections past and contemporary writes, more elusively, "It is now commonly understood that the success of the prescientific *Literary Digest* poll from 1908 through 1932 rested on the absence of any autonomous class cleavage in electoral politics, just as the poll's disastrous failure in 1936 reflected the emergence of this fatally decisive variable" (Burnham 1987: 119). Implicit in this argument is that the *Digest* sample was biased toward the more affluent, but because "social class" was not a variable that had an effect on voting preference before 1936, one could rely on any one class to predict presidential elections. It will be shown later that there is evidence to contradict this assertion.

The author of a little book on the misuse of statistics that has enjoyed a great deal of popularity writes: "People who could afford telephones and magazine subscriptions in 1936 were not a cross section of voters. Economically they were a special kind of people, a sample biased because it was loaded with what turned out to be Republican voters" (Huff 1954: 20). Other examples can be found. For instance, a well-respected textbook on survey research methods states: "The *Digest* sample was drawn from telephone directories and automobile registration lists, a sampling procedure that . . . did not provide a representative cross section of American voters" (Babbie 1990: 67). One can also refer to a current statistics manual and read: "[The] *Literary Digest* magazine tried to predict the outcome [of the presidential election] by mailing 10 million questionnaires to people selected from three sources: the subscription list for the magazine, telephone directories, and automobile registration records" (Bernstein and Bernstein 1999: 63). The

same picture, although more nuanced, is presented in another text on sample surveys; explicitly using Gallup as a reference, the author writes: "One of the reasons that [the *Digest* poll] pointed to the wrong candidate was that the voters selected in the sample were, for the most part, telephone subscribers and owners of automobiles" (Raj 1972: 331).[13] This view of the cause of the *Digest* poll's blunder has also propagated itself beyond our borders (Desrosières 1998: 233; Gidengil 1992: 244; Moon 1999: 10–12; Ozouf 1963: 4; Robinson 1999: 40; Weiner 1976–77: 675).

This review is by no means exhaustive (see also, among many others, Bradburn and Sudman 1988: 19; Burner and Rosenfield 2003: 409; Erikson 1976: 26; Erikson and Tedin 1981: 953;[14] Field 1983: 198–99; Herbst 1993: 70; Igo 2006: 111; Kitchens 1987: 8; Ladd 1985: 332; Oberschall 2008: 87; Weisberg et al. 1998: 42). But it is highly suggestive: it clearly shows that the conventional explanation is firmly entrenched in the scholarly community and, by extension, in the interested public at large (e.g., the press), and it persists even after evidence has surfaced that contradicts it (Squire 1988). In her history of survey research in the United States, Jean M. Converse (1987: 120) states: "Gallup's theory of the *Digest*'s failure has not been improved upon in any major way." Later it will become evident, through a poll conducted by Gallup's own organization, that his "theory" is not supported by the data.[15]

Doubting the Conventional Explanation: A Skeptic and Empirical Data

The conventional explanation was never seriously challenged until 1976. This happened in a short article titled "The *Literary Digest* Poll: Making of a Statistical Myth" published in the *American Statistician*. The author, Maurice Bryson (1976: 184) found it "inherently implausible." In his view, it could not explain the huge error made by the *Digest*; that error and the low response rate, Bryson (ibid.: 185) argued, pointed to "*voluntary response*," and thereby nonresponse bias, as the more likely candidate for the poll's blunder. But he presented no evidence for his argument.

Surely, if there were data that indicated how telephone and car owners voted in the 1936 election, we could move away from the realm of conjecture (Gallup's and Bryson's). If we found that telephone and car owners voted massively for Landon, the conventional explanation would be vindicated; if, on the other hand, we found the opposite, then it could be dis-

carded and another explanation for the poll's blunder could be sought (perhaps Bryson's).

Actually, this evidence did exist. It had been gathered by a Gallup poll in May 1937. How did that come about? By this time, the AIPO was conducting polls on a monthly basis, and more than five dozen newspapers nationwide were subscribing to its service and publishing the results (News–Week 1936: 14). Don Cahalan, a student at the University of Iowa, was working toward a master's degree in psychology. The focus of his research was to identify the factors affecting the validity of mail-ballot polls (Cahalan 1989: 132). Cahalan asked the AIPO to include questions about the 1936 *Digest* poll in one of its surveys. His thesis director, Norman C. Meier, had been one of Gallup's professors at Iowa (Ohmer 2006: 82). The questions were added to a Gallup poll that dealt with current issues and was fielded between May 19 and May 24, 1937. Aside from asking whom respondents had voted for in November 1936, the questionnaire included three questions about the *Digest* poll:

· Did you receive a Literary Digest straw vote ballot in the Presidential campaign last fall?
· Did you send it in?
· Did you change your mind regarding the candidate between the time you sent it in and the election?
(Roper Center for Public Opinion Research 2003a: 3)

The AIPO survey also collected demographic information, such as occupation, age, social/income status ("poor," "average income," etc.), ownership of a telephone and automobile, gender, and "race" (white or "colored").

Cahalan had also conducted his own survey in Cedar Rapids, Iowa, a "representative community" (Cahalan and Meier 1939: 4), between December 1936 and February 1937 that was based on telephone interviews of a systematic random sample of $n = 693$. The survey achieved a response rate of 80 percent.[16] Cahalan and Meier had used a sampling frame of more than 10,000 names from the list the *Digest* had put together from the telephone book (Cahalan 1989: 130).

The results of both surveys were analyzed by Cahalan and Meier in the February 1939 issue of the *Psychological Record*. The authors noted that the reported response rate to the *Digest* poll from participants in both the Cedar Rapids survey (59 percent) and the AIPO poll (61 percent) was much higher than the actual one of 24 percent (Cahalan and Meier 1939: 5–6). They

reported also that both surveys indicated that the *Digest* poll was biased on a number of variables: age, sex,[17] and party preference. Regarding the latter, they failed to state that the difference, in Cedar Rapids, between Roosevelt (47 percent) and Landon (53 percent) was not statistically significant and was consistent with the election results in that city: each candidate received an equal share of the vote (Cahalan 1989: 131). Thus on this evidence, it was incorrect of Cahalan and Meier (1939: 8) to conclude that the original *Digest* list for Cedar Rapids suffered from "over-Republicanism." They observed the same flaw in the *Digest* (national) sample, as described by the AIPO poll, but they neglected to point out that it favored Roosevelt (56 percent) over Landon (44 percent).[18] However, they did emphasize that this error was "minor" compared with "factors related to ballot return" (ibid.). Indeed, their results show that in the AIPO survey nearly 75 percent of Landon sup- porters in the *Digest* poll reported returning their straw ballots, compared with only 57 percent of Roosevelt backers. The same happened in the Cedar Rapids survey: Landon's followers were twice as likely to return their straw ballots as Roosevelt's (ibid.: 6).

These results plainly contradicted the conventional explanation that was advertised by the protagonists at the time (e.g., Robinson 1939: 3) and has persisted until this day. According to these data, what sank the *Digest* poll was nonresponse bias. But the most glaring omission in this 1939 article was the fact that the authors never discussed how telephone and automobile owners voted, even though this information was available from the AIPO survey. This is an amazing oversight, since the reliance on telephone directo- ries and car registration lists was the defining characteristic of the *Digest* poll. Moreover, Cahalan and Meier say nothing of the fact that the poll in Cedar Rapids relied entirely on telephone subscribers, a source that does not seem to have introduced a political party bias.

Clearly, Cahalan and Meier's article had little impact on the scholarly community and beyond.[19] It was not even cited by Peverill Squire in his 1988 article in the *Public Opinion Quarterly*, the flagship journal of survey research practitioners. The irony is that, at the time of his study of the 1936 poll's fail- ure, Squire was an assistant professor of political science at the same uni- versity where Cahalan had done his research. Squire used the same 1937 AIPO poll that Cahalan had been instrumental in designing but appears to have been unaware that it had already been reported on.[20] Cahalan (1989) remarked on the irony a year later in the *Public Opinion Quarterly*.[21]

Like the previous authors, Squire (1988: 129) noted that the poll had an overrepresentation of people who claimed to have returned their *Digest* ballots, but he also pointed out that it overestimated the percentage received by the winning candidate. His conclusions, however, were similar to what was apparent from Cahalan and Meier's article. Squire (ibid.: 130) stated that "if everyone who had received a [*Digest*] ballot had returned it the results would have, at least, correctly predicted Roosevelt a winner." It was clear to him that "nonresponse bias contributed greatly to the failure of the *Literary Digest* to correctly call the winner" (ibid.: 131). Yet his overall conclusion was somewhat muddled. Thus, after acknowledging the importance of nonresponse bias, he stated: "But, more importantly, the initial sample was flawed" (ibid.). Why "more importantly," when he had just demonstrated that the *Digest* would have called the correct winner if everyone had returned his or her straw ballot? Perhaps it is because he attempts, using a "rough calculation" (ibid.), to rank in terms of importance the two components of the overall error: sample bias and nonresponse bias, which, he estimates, account for 11 points and 7 points, respectively (ibid.).[22]

The main difference between Squire's article and Cahalan and Meier's, which has a direct bearing on the issue of sample bias and thus on the conventional explanation, is that Squire looked at the way telephone and automobile owners voted in 1936. He noted that the AIPO poll "overrepresents the percentage of people who had a telephone or a car" (ibid.: 129). The results, nevertheless, showed that "owners of only an automobile or a telephone were less supportive of Roosevelt than those who did not have either, but they were still strongly for him. Even respondents who had both a car and an automobile were for the president" (ibid.). The data presented by Squire (ibid.: 130) indicated that, as a group, 61 percent of telephone and car owners favored Roosevelt over Landon. Here was concrete evidence that the conventional explanation, a mere conjecture, was wrong. If relying on telephone directories and car registration lists was not the problem, since owners of these items favored Roosevelt, then the cause of the *Digest* poll's blunder must lie elsewhere.

But can these poll results be relied on to reach a valid conclusion? After all, the AIPO survey was based on quota sampling, long frowned on by survey practitioners in the United States; in addition, this particular poll overrepresents Roosevelt voters as well as owners of telephones and automobiles. Although, as Squire (ibid.: 129) pointed out, "This survey is the best avail-

able tool for determining why the *Literary Digest* poll failed," the data collected should not be analyzed in raw form. In the next section I apply statistical techniques commonly used by survey data analysts nowadays to adjust the data to known population values. Neither Cahalan and Meier nor Squire attempted such an adjustment; hence the limitation of their conclusions.

A Reanalysis of the May 1937 Gallup Poll

Although Gallup polls of the 1930s and 1940s have been used by many scholars (e.g., Andersen 1979; Baum and Kernell 2001; Ladd and Hadley 1975) and although pollsters in those days referred to them as "scientific," they are based on a sampling technique (quota control) that does not meet contemporary canons of sound statistical practice (probability sampling). Despite this major flaw, Berinsky (2006: 509) believes that these data have "inherent value" from which inferences can be drawn if the proper statistical methods are used. Following in Berinsky's footsteps, I rely on weighting techniques that have long been used by survey data analysts.[23] The general approach adopted is known as *poststratification weighting*. It is commonly applied in surveys to reduce any bias resulting from nonresponse and noncoverage; the goal is to adjust the data so that they conform to known population values and "to improve the precision of survey estimates" (Kalton and Flores-Cervantes 2003: 81, 82). This section focuses on two primary issues: how telephone and car owners voted and what the candidate preferences were among the original *Digest* list of 10 million.

In the first case, our reference population is the electorate: the population of voters (45.6 million in 1936). Here there is a problem, because in those days the census did not yet conduct the Current Population Survey, with its election-year November supplement on registration and voting, nor did the American National Election Studies or exit polls exist. In other words, not all the variables in the analysis have "known population values." I have mentioned that the 1937 AIPO poll overestimates Roosevelt's share of the total popular vote (65.8 percent instead of 60.8 percent) and overrepresents owners of telephones and owners of automobiles (51.7 percent instead of 33.1 percent and 61.0 percent instead of 57.0 percent, respectively).[24] But this overrepresentation of car and telephone owners is assessed with respect to values in the *general* population, not the *voting* population. Thus the results that I am about to present are based on (1) the assumption that the distri-

bution of these consumer items are similar in both populations and (2) the known distribution of votes for each presidential candidate. Further, these results are obtained using a weighting method referred to as *raking* or *iterative rim weighting*.[25] The purpose is to adjust the frequencies in the cells, of which table 1 has 12 (3 rows × 4 columns), so that the marginal totals (row totals and column totals) will conform to known population values. This technique is used when the distribution of each variable (item ownership and voting preference) is known but their joint distribution is unknown.

Table 1 shows that Roosevelt receives substantial majorities among two of the three categories of car and telephone owners. Only among the owners of both consumer goods does Landon have a lead.[26] This is also the only major difference between the weighted and unweighted data. The other results are all in the same direction, although, unsurprisingly, Roosevelt's share of the vote is less than the original (unweighted) data imply. The main conclusion is that, as a group, a majority of telephone and automobile owners (53.9 percent) favored the incumbent president: this contradicts the conventional explanation and confirms the results presented by Squire (1988: 130). Therefore reliance on telephone directories and automobile registration lists cannot be the reason that the 1936 *Digest* presidential poll failed to forecast a Roosevelt victory.

I focus now on the *Digest* sample based on the outcome of the AIPO's 1937 poll. Table 2 indicates that Landon is ahead of Roosevelt among *Digest* poll respondents but behind him among nonrespondents.[27] Finally, the results for the overall sample (respondents and nonrespondents combined), that is, the original list of 10 million, show that Roosevelt is backed by a majority. The reference population for this analysis is the original list (10 million) that the *Digest* used to send out its straw ballots for the 1936 presidential election. Given the available information regarding the *Digest* poll, there are clear anomalies with the sample from the AIPO survey. First, the response rate to the *Digest* poll, as reported by the AIPO, is 64 percent, when in fact it should be 24 percent. In other words, the Gallup survey oversampled *Digest* respondents.[28] The main task is to estimate how the original *Digest* sample (the list of 10 million) would have voted in the straw poll; this excess of respondents could bias the results, since, as table 2 shows, candidate preferences for respondents and nonrespondents are dissimilar. In other words, if the raw results are not corrected, in effect too much weight is given to respondents and not enough to nonrespondents. Furthermore, the results

Table 1 Candidate preference by consumer items (in percent)

Candidate	Car and telephone		Car only		Telephone only		Neither		Total	
Roosevelt	*46.4*	*(54.9)*	*59.2*	*(67.6)*	*61.9*	*(69.5)*	*72.5*	*(79.1)*	**60.8**	(65.8)
Landon	*52.5*	*(44.5)*	*36.7*	*(30.0)*	*37.4*	*(30.1)*	*24.6*	*(19.2)*	**36.5**	(32.9)
Other	*1.0*	*(0.6)*	*4.2*	*(2.5)*	*0.7*	*(1.7)*	*3.0*	*(1.7)*	**2.7**	(1.3)
Total[a]	**27.3**	(41.4)	**29.7**	(19.6)	**5.8**	(10.3)	**37.2**	(28.7)	2,285[b]	

Source: Roper Center for Public Opinion Research 2003b.
Note: The original unweighted results are in parentheses. The weighted results are italicized. The bolded numbers for the row totals are known population values. The bolded numbers in the last row of the table are estimated population values based on known population values for car ownership and telephone ownership. Some percentages do not sum to 100 due to rounding.
[a] Row percentages.
[b] Total sample size.
The statistical test results of the two-party differences are car and telephone, $p > .10$; car only, $p < .001$; telephone only, $p < .01$; neither, $p < .0001$.

Table 2 Candidate preference of *Digest* poll respondents and nonrespondents (in percent)

Candidate	*Digest* poll respondent				Total	
	Yes		No			
Roosevelt	**42.9**	(48.5)	*60.9*	(66.1)	*56.6*	(54.8)
Landon	**57.1**	(51.5)	*39.1*	(33.9)	*43.4*	(45.2)
Total[a]	**23.8**	(63.7)	**76.2**	(36.3)	764[b]	

Source: Roper Center for Public Opinion Research 2003b.

Note: This table indicates voting preference *at the time of* the *Digest* poll; that is, it accounts for those who remembered changing their minds ($n = 26$); those who claimed not to have changed their minds ($n = 433$); those who said that they did not remember ($n = 13$); and those who gave no answer ($n = 292$). The original unweighted results are in parentheses ($n_{11} = 236$; $n_{12} = 183$; $n_{21} = 251$; $n_{22} = 94$). The weighted results are italicized ($n_{11} = 78$; $n_{12} = 355$; $n_{21} = 104$; $n_{22} = 228$). (Weighted cell frequencies do not sum to total due to rounding.) The bolded numbers are known (population) values.

[a] Row percentages.

[b] Total sample size. It is made up of AIPO poll respondents who report having received a *Digest* ballot. Of those, *respondents* are individuals who claim to have returned their straw ballots ($n = 487$), and *nonrespondents* are individuals who said that they did not return theirs ($n = 246$) or did not remember returning it ($n = 31$).

on candidate preference among respondents are at variance with the actual results of the *Digest* poll: in the latter Roosevelt received 42.9 percent of the two-party vote; here he gets 48.5 percent of the straw ballots.

As with the earlier analysis concerning car and phone owners, the raw results must be weighted so that they conform to known population values. But an awkward situation develops: the known information is a combination of cell values (voting preference of *Digest* poll respondents) and marginal totals (the overall response rate of the *Digest* poll). In other words, half the cell frequencies in the two-by-two table (table 2) have to be estimated, but *raking* cannot be used because some marginal totals (the row totals) are missing. The column totals can be adjusted to reflect the true response rate to the *Digest* poll. Then, *cell weighting* can be used for the cells made up of *respondents* to the *Digest* poll, but not for the other half—the cells that correspond to *nonrespondents*.[29] For those cells, the compromise solution, in the absence of better information, is to use the same correction factors that are applied to the respondents' cells and then adjust the results to conform to the corrected column total for nonrespondents.

On the basis of these data, the following conclusions regarding the failure of the 1936 *Digest* poll can be reached:

1. Had everybody on the original *Digest* list returned his or her straw ballot, the magazine would have been in a position to forecast the correct winner of the election: Roosevelt.

2. Poll respondents and nonrespondents favored opposite candidates: while three-fifths of nonrespondents voted for the incumbent, only two-fifths of respondents did.

3. Landon's supporters were much more likely to return their straw ballots than Roosevelt's: nearly one-third versus only one-fifth.

4. The bulk of the forecast error made by the *Digest* can be attributed to nonresponse bias. Roosevelt's actual share of the two-party vote was 62.5 percent; his share as forecast by the *Digest* was 42.9 percent: a difference of 19.6 points—that is the extent of the total error made by the *Digest*. Referring back to table 2, this error can be partitioned into two additive components: the portion due to sample bias ($5.9 = 62.5 - 56.6$) and the portion resulting from nonresponse bias ($13.7 = 56.6 - 42.9$).

I will end this section by illustrating how the *Digest* poll gave the opportunity for Landon supporters to vent their displeasure with the New Deal administration, irrespective of sociodemographic background.[30]

It has been said that the 1936 contest generated "the bitterest campaign since 1896" (Hofstadter et al. 1959: 527).[31] The New Deal, which had ruffled many feathers, was accused, among other things, of being "socialistic," the ultimate anathema in the political culture of America. Feelings ran high in certain quarters of the American electorate that "America [was] in peril" as a result of the administration's policies. It is no wonder, then, that highly motivated individuals seized on the *Digest* poll to register their discontent with the direction in which the administration was taking the country. These respondents caused the downfall of the poll and, ultimately, of the *Digest* itself.

Table 3 shows that the level of participation was higher among Landon's partisans than among Roosevelt's supporters regardless of the ownership category. Ironically, the highest turnout rate in favor of Landon was among the "have-nots" (40.5 percent), followed by the owners of both commodities (34.5 percent). The same pattern is present in table 4, which describes the occupational status of the *Digest* sample: in all groups, even the unemployed, Landon supporters sent back their straw ballots at a higher rate. In table 5 those who received the *Digest* ballot are classified into age groups.

Table 3   Response rate by ownership and candidate preference (in percent)

Consumer item	Candidate	Returned straw ballot	
		Yes	No
Car and telephone** (*n* = 475)	Roosevelt	16.9	83.1
	Landon	34.5	65.5
Car only* (*n* = 134)	Roosevelt	13.8	86.2
	Landon	17.4	82.6
Telephone only* (*n* = 79)	Roosevelt	22.8	77.2
	Landon	30.3	69.7
Neither* (*n* = 77)	Roosevelt	24.5	75.5
	Landon	40.5	59.5

Source: Roper Center for Public Opinion Research 2003b.
Note: *Digest* subsample (*n* = 764). Row percentages.
**$p < .001$. *not significant.

One exception to the general pattern already observed occurs among those 21 to 24 years old: in that cohort, unlike all the others, Roosevelt supporters returned their straw ballots at a higher rate (18.1 percent) than Landon followers (7.8 percent).

It is clear from these results that Landon supporters were far more motivated to participate in the *Digest* poll than backers of the incumbent president, and this motivation was somewhat more pronounced among higher-status occupational groups (professional, business, and skilled worker) and older individuals than among the others, but ownership of a car or a telephone does not seem to have made much difference. Thus a combative opposition to Roosevelt and the policies he represented seems to have been, for respondents, the central trigger to their participation in the poll, which provided a venue for these people to express a strong dislike of the New Deal.[32]

Discussion

The results of this study demonstrate that the conventional explanation as an account of the failure of the 1936 *Digest* poll is indeed a "myth" (Bryson 1976: 184). However, this research is by no means definitive, nor can it be: we simply do not have the data to reach an unassailable conclusion. The only sure way to make a final determination on this issue, according to contem-

Table 4   Response rate by occupational status and candidate preference (in percent)

Occupational status	Candidate	Returned straw ballot	
		Yes	No
Professional*** (n = 77)	Roosevelt	16.2	83.8
	Landon	42.8	57.2
Business** (n = 123)	Roosevelt	19.3	80.7
	Landon	34.7	65.3
Skilled worker*** (n = 173)	Roosevelt	20.8	79.2
	Landon	37.4	62.6
Unskilled worker* (n = 139)	Roosevelt	19.9	80.1
	Landon	28.9	71.1
Farm* (n = 169)	Roosevelt	14.6	85.4
	Landon	19.2	80.8
Other* (n = 64)	Roosevelt	17.3	82.7
	Landon	20.8	79.2
Unemployed[a]*** (n = 19)	Roosevelt	13.0	87.0
	Landon	100.0	0.0

Source: Roper Center for Public Opinion Research 2003b.
Note: *Digest* subsample (n = 764). Row percentages.
[a] Fisher's exact test was used.
***p < .05; **p < .10. *not significant.

porary canons of scientific practice, would be to select a probability sample from the *Digest* list of 10 million, just as Cahalan did for Cedar Rapids; otherwise the empirical evidence presented and discussed here is the best we have. As Squire (1988: 129) himself stated, the evidence is "less than perfect." But how serious are these imperfections?

I have already mentioned the issue of quota-controlled sampling. It is especially a problem, since, in those early days of "scientific" polling, interviewers were largely left to their own devices on how to go about filling their quotas.[33] We know that Gallup shaped his samples to be a miniature of the voting population; to that end he deliberately undersampled certain individuals, such as women, blacks, and the poor. It is easy, when the information is available, to compare AIPO samples to census data for the general population. The AIPO's (1938: 9) samples were selected based on quotas for six variables: proportion of voters represented by each state, area of residence (urban and rural), gender, age, "income," and party preference.[34] The first

and the last are the only two variables that can be compared to the actual distribution of voters. For this sample, the representation of voters by state is fairly close: it overrepresents both the South by 1.4 points and New England by 0.5 point; other regions—Middle Atlantic, Midwest/Central, and West—are underrepresented by 1.7, 3.5, and 3.2 points, respectively. Given that this sample was not designed specifically to forecast an election, the results are remarkably accurate.[35] The breakdown by area of residence is less accurate, but the comparison is not based on the same populations: the AIPO sample (urban: 64.7 percent; rural nonfarm: 15.7 percent; rural farm: 19.6 percent) is intended to represent voters; the census data (urban: 56.4 percent; rural nonfarm: 20.1 percent; rural farm: 23.5 percent) is for the entire population (US Bureau of the Census 1997: Series A-73).[36] The Gallup sample would seem to overrepresent the urban portion of the population. The same comparative asymmetry applies to gender and age. In the general population, the proportion of males (50.3 percent) and females (49.7 percent) aged 21 and older is about even (US Bureau of the Census 1931: 20; 1943: 38); in the Gallup sample, males (67.1 percent) outnumber females (32.9 percent) two to one. It would appear that the 1937 AIPO poll substantially undersampled women. But again, the population of interest is the electorate, not the general population: it is generally believed that women in those days did not participate in elections as much as men—to what extent it is difficult to ascertain.[37] If the AIPO's sample is gender biased, does it really matter? It does if men and women voted differently in the presidential election. There appears to be no scholarly consensus on these issues.[38] In the 1937 sample, male voters seem to favor Roosevelt slightly more (62.8 percent) than female voters (57.7 percent); although the difference is significant, it is not large, and both groups favor the same candidate.[39] If I compare the distribution of age categories (see table 5) of the AIPO sample to census data (US Bureau of the Census 1997: Series 31–37, 41), the two youngest (21–24 and 25–34) and the oldest (55+) groups are underrepresented by 5.0, 0.8, and 4.3 points, respectively.[40] As a result, the other two age categories (35–44 and 45–54) are overrepresented by 5.9 and 4.1 points, respectively.[41] This is not surprising, given the oversampling of better-off folks—those with a car or a telephone or both (see table 1). Moreover, interviewers contributed to this process through their own social biases. As one contemporary observer remarked: "The upward bias probably results in large part from the reluctance of middle-class interviewers to approach the lowest income groups" (Rugg 1947: 146). But how relevant is

Table 5   Response rate by age and candidate preference (in percent)

Age category (years)	Candidate	Returned straw ballot	
		Yes	No
21–24* (n = 34)	Roosevelt	18.1	81.9
	Landon	7.8	92.2
25–34*** (n = 153)	Roosevelt	20.4	79.6
	Landon	41.3	58.7
35–44* (n = 222)	Roosevelt	21.1	78.9
	Landon	27.3	72.7
45–54** (n = 212)	Roosevelt	16.6	83.4
	Landon	27.3	72.7
55+**** (n = 139)	Roosevelt	11.1	88.9
	Landon	36.7	63.3

Source: Roper Center for Public Opinion Research 2003b.
Note: *Digest* subsample (*n* = 760). Row percentages.
****$p < .001$. ***$p < .01$. **$p < .10$. *not significant.

this "bias toward the upper economic levels" (ibid.) in the AIPO samples to the problem at hand? Under the conventional explanation assumption, one would expect such a biased sample to display a distinct preference for the Republican candidate. Instead, table 1 shows that the "haves" (owners of cars and telephones) favored the incumbent president. However, this conclusion is not valid if based on the raw data, because the sample overrepresents the proportion of voters who favored him. In other words, the original sample has too many "haves" and too many supporters of the president. This jeopardizes any inference regarding the candidate preference of car and telephone owners. Indeed, although the overall excess in favor of Roosevelt (5.0 percent = 65.8 percent − 60.8 percent) is known, the overestimate in each cell is not. Before I discuss this issue further, I will try to figure out the process that gave rise to this situation.

Squire (1988: 129) explains this discrepancy as follows: "Like almost all postelection surveys, it overestimates the vote for the winning candidate." Is that right? What Squire is suggesting is that a number of poll respondents who voted for Landon in the November presidential election misreported their voting preference for the poll. This can be done in three ways: deceptively (the respondents remember that they voted for Landon but choose to report having voted for Roosevelt), deliberately (they cannot remember their

choice but do not want to admit it and decide to report that they voted for the winner), or involuntarily (the respondents *believe* that they voted for the candidate they report on but actually voted for the opponent).[42] According to research done on vote choice, deception is unlikely; memory lapse is the more common cause of misreports (Wright 1993: 292). The same study indicates that the rate of misreport in presidential elections is low (ibid.: 294, 310) compared to other electoral contests (Senate, House, governor). Thus it would seem that what occurred with the 1937 AIPO poll is not so much a misreport of candidate choice in favor of the winner as an oversampling of Roosevelt supporters.[43] Since the poll was not designed to predict an election, it is possible that interviewers were not as constrained to produce a balanced sample along party lines as they would have been in a preelection poll.[44]

The poststratification weighting strategy adopted in this article, which is commonly used in probability samples to compensate for nonresponse and noncoverage, rests on a leap of faith: that respondents are representative of nonrespondents. In the case of Gallup's samples, were those interviewed similar to those of the same background who were passed over by the interviewers? Were the owners of cars and telephones interviewed by the AIPO representative of voters of the same profile? Were those in the Gallup sample who received the *Digest* ballot representative of those on the *Digest*'s original list? We will never know, but I made the assumption that they are comparable and used weighting techniques to correct for obvious biases.[45]

To resolve the issue of working with a sample that overrepresents both the economically better-off and the supporters of the incumbent, I relied on *iterative rim weighting*. This statistical technique allows the user to estimate the counts in the cells of the table that cross-classifies ownership of a car or a telephone with voting preference (table 1) so that cell frequencies will add up to the marginal totals (known population values). Squire's analysis of the same data made no attempt to adjust ownership; as for voting preference, he applied the marginal overestimate of Roosevelt (5 points) and underestimate of Landon (4 points) to each corresponding cell in the table (Squire 1988: 130). This assumes that the over- and underselection of respondents was the same for each category of ownership—a most unrealistic assumption. Results in table 1 show that the oversampling of Roosevelt voters varied from 8.5 points for owners of both a car and a telephone to 6.6 points among the "have-nots," while the underrepresentation of Landon voters varied from

8.0 to 5.4 points, respectively. My conclusions, however, are not fundamentally different from those reached by Squire in that a majority of automobile and telephone owners voted for Roosevelt (53.9 percent, 95 percent confidence interval: 51.3, 56.5).

The same lack of statistical adjustment pervades the analysis of the *Digest* poll as described by the Gallup sample: neither Cahalan and Meier (1939) nor Squire (1988) used available auxiliary information to base their analyses on sound statistical ground. Moreover, Squire's (ibid.: 131) attempt to partition the overall forecast error made by the *Digest* between what can be accounted for by sample bias and what is related to nonresponse bias is flawed: he computes this breakdown without taking into account the fact that respondents are overrepresented. Since respondents and nonrespondents had very different candidate preferences (table 2), Squire's estimate is unsound.

Conclusions

This study had two related objectives: test empirically the validity of the conventional explanation and, if it proved wrong, find another explanation for the *Digest* poll's massive forecast error.

The conventional explanation, which assumes that, because the *Digest* relied mainly on telephone books and car registration lists as its sampling frame, the poll's original sample had to be biased in favor of the Republican challenger, is mistaken. The data indicate that if the *Digest* had relied entirely on these lists, it would have forecast Roosevelt as winning the election. Since the sampling frame was not to blame, another explanation was required. Based on a sample of individuals who reported receiving a ballot from the *Digest*, the data show that nonrespondents, unlike respondents, favored Roosevelt heavily and that a majority of the original list also supported him. In other words, nonresponse bias was the principal culprit for the poll's failure: if all the straw ballots from the initial *Digest* sample of 10 million had been returned instead of only 24 percent of them, the *Digest* poll not only would have called the correct winner but would have done probably as well as two out of the three new "scientific" pollsters (Crossley and Gallup: see Gallup 1972b: 215; Katz and Cantril 1937: 164).[46] Roosevelt, who got 62.5 percent of the two-party vote, was predicted by the Gallup poll to receive 55.7 percent (Gallup 1972b: 215), a 6.8 point underestimate; the

data presented here (table 2) show that 56.6 percent (95 percent confidence interval: 53.1, 60.1) of the *Digest*'s original list supported him—a 5.9 point underestimate.

Despite the evidence uncovered by Squire, which should have caught the attention of social science researchers and other writers who refer to the *Digest* poll, the conventional explanation, "accepted by nearly everybody simply because nearly everybody repeats" it (Bryson 1977: 181), does not seem to want to go away (e.g., among others, Fallacy Files n.d.; Kircher 2007; Saxon 2000). Like other myths, such as Isaac Newton's apple, it has acquired a life of its own; its popularity will be hard to dispel, because it has this quality so well captured by the Italian expression "Se non è vero è ben trovato" (Even if it's not true, it's still a good story). The more this "explanation" is repeated, the more its factual status is enhanced. It has considerable intuitive appeal: it is easily remembered and has a distinct ring of truth. It is also important that its pedigree can be traced back to such an authority as Gallup: that makes it almost ironclad. From the outset, however, it was a conjecture and required empirical evidence to be confirmed or rejected. The May 1937 AIPO poll has provided the data to test this hypothesis.[47] I found that the evidence contradicted the assumptions associated with the conventional explanation.

Are the limited and imperfect data presented in this study preferable to mere conjecture? Aside from the 1937 AIPO poll, we have the *Digest* poll's results from three cities (Allentown, Scranton, and Chicago) in which only registered voters were sampled: all suffered from low response rates, and all favored Landon (like the national sample). There is also the probability-based survey conducted by Cahalan shortly after the election in Cedar Rapids: there again, Landon was favored by respondents. Furthermore, the data show that nonrespondents preferred Roosevelt and that the list used by the *Digest* for that city was unbiased. What is remarkable about these sets of results is their consistency: they all show that nonresponse bias was the primary cause of the *Digest* poll's massive error. This is what one might call a "confluence of evidence" (Ziesel 1985: 252).

For social scientists who rely on sample surveys as their source of data, the lesson from the *Digest* is that any poll or survey that has a low response rate is probably biased. It is ironic that, in this age of the Worldwide Web, many polling houses rely on huge Internet "panels" that bring to mind the size of the original list of the *Digest*.[48] These panels, just like the *Digest* list,

suffer from noncoverage (a nontrivial portion of US households is still not connected to the Internet). Organizations that use a combination of probability sampling and the Web to collect their data are fraught with low response rates. Of course, nowadays they have at their disposal a wide array of statistical tools[49] and population figures to adjust the data once the field period of the survey is over.

The second lesson that can be derived from this study merely reiterates what Berinsky (2006) has already suggested: social scientists who use poll results from the 1930s and 1940s are well advised to adjust the data to known population parameters when these are available from the census.

In conclusion, the available empirical evidence I have reviewed suggests that the principal cause of the 1936 *Literary Digest* poll fiasco was most probably nonresponse bias: the result of a low response rate (24 percent) combined with a very large difference in candidate preference between *Digest* poll respondents, who heavily favored Alf Landon, the Republican challenger, and nonrespondents, who in turn strongly supported the incumbent president, Franklin D. Roosevelt.

Notes

1   These lists are referred to as the *sampling frame*, because they are the list of population elements from which the sample is drawn.

2   Gallup is referring to the quota method of sampling, *not* probability sampling.

3   For a history of this concept, see Kruskal and Mosteller 1979a, 1979b, 1979c, 1980.

4   "Dr. Gallup seeks to sample only voters" (Stephan 1939: 347).

5   Of the three "pioneers," Ohmer (2006: 4) writes: "Gallup publicized these new [polling] methods more vigorously. His interviews in popular magazines and frequent speeches to business and community groups made him the symbol for these new 'scientific' techniques."

6   Nonresponse bias is the difference between the way respondents and nonrespondents to a poll or survey answer a given question item—here their choice of presidential candidates.

7   Since 1924 the *Digest* poll had correctly predicted the outcome of presidential elections (*Literary Digest* 1936c: 6).

8   The *Digest* did not use its own subscriber list as part of the sample (*Literary Digest* 1928: 7).

9   Unless otherwise indicated, all the quotes that follow are taken from this issue of the *Literary Digest* 1936d: 7–8.

10   However, the editor, Wilfred J. Funk, was quoted in the *New York Times* (1936) as saying that "we may not have reached a cross-section of the population in distributing our ballots." Ironically, Gallup (1938: 139) admitted the same as one factor explaining his underestimate of Roosevelt's share of the vote: "We under-estimated lower income groups in our sample."

11   In the same article, a few pages later, Stephan (1939: 348) writes: "Experience in questionnaire inquiries has shown that persons who return questionnaires differ in some important characteristics from those who do not and consequently a sample composed of such persons may be badly biassed [*sic*]. The unfortunate experience of the *Literary Digest* was *principally due* to this factor" (emphasis added). This directly contradicts the statement quoted above.

12   As one author states, Gallup was the main promoter of "the popular notion that the [*Digest*] poll failed because of biases in the sample" (Grier 2002: 32).

13   The author relies on Gallup and Rae 1940.

14   Erikson and Tedin (1981: 953) add an original twist to the conventional explanation, although they offer no empirical evidence for it. They argue that the 1936 *Digest* list was made even more biased because its reduction from 20 million to 10 million was done at "the expense of the less affluent."

15   In fairness to Converse, her book was published a year before Squire 1988. However, in Converse 2009, a reprint that includes a new introduction, no change regarding this issue is mentioned.

16   Eight percent refused, and another 12 percent could not be reached. The effective sample size was $n = 554$.

17   They compared the distribution of the demographic variables in the *Digest* list to census data of the *general* population. The authors do not mention that this is not an entirely valid comparison, since the *Digest* list was purported to represent the *voting* population.

18   See table 1 in Cahalan and Meir 1939: 6. Roosevelt received 63 percent of the two-party vote, so there is a 7-point bias $(63 - 56)$ in the original *Digest* list based on the raw results of the Gallup poll.

19   A search on Google Scholar (May 22, 2010) reveals 10 citations of this work: 3 in the 1940s, 1 in the 1950s, 2 in the 1970s, 1 in the 1980s, 1 in the 1990s, and 2 in the first decade of the 2000s. Only one of these citations (Cahalan 1989) discusses the cause of the *Digest*'s poll blunder.

20   In the article's abstract Squire (1988: 125) states: "No empirical research has been conducted to determine why the poll failed." Later in the article he writes that the AIPO poll "has never, to my knowledge, been exploited" (ibid.: 128).

21   In that article Cahalan (1989: 129) states that his "conclusions in 1938 were the same as Squire's in 1988." He comments mostly on his Cedar Rapids study but never

directly addresses the issue of the *Digest*'s sampling frame (telephone directories and automobile registration lists).

22   This estimate is incorrect, because it is based on a sample that gives too much weight to respondents heavily in favor of Landon. The next section shows that Squire's conclusion on this matter is wrong.

23   These methods have been around since the 1940s. See Deming 1964 [1943]; Deming and Stephan 1940; and Stephan 1942.

24   The official reference for the May 1937 poll is USAIPO1937-0083 (Roper Center for Public Opinion Research 2003b). I assume that each poll participant represents one household. For the telephone percentage: US Bureau of the Census 1997: Series R3. For automobiles: Fischer 1992: 102; US Bureau of the Census 1997: Series Q153.

25   For a description of the raking process, see Berinsky 2006: 513 and Kalton and Flores-Cervantes 2003: 86. It is iterative because it usually takes several iterations (of computations) before the sum of the cells (the intersections of rows and columns) converge to the marginal totals that are the targets.

26   The difference between Landon and Roosevelt is not statistically significant in this category (see table 1 note).

27   For the purpose of this analysis, only the two-party vote is relevant.

28   This is not surprising, since the survey also oversampled car and telephone owners.

29   Cell weighting is used when the joint distribution of the variables is known. Here the joint distribution for one value of one dichotomous variable (respondents) and for two values of the other variable (candidate choice: Roosevelt or Landon) are known.

30   The data in tables 3–5 are weighted.

31   My remarks are based on these authors' assessment of the 1936 election and the New Deal.

32   The *Digest* was not the only straw poll that favored Landon: so did the *Farm Journal* and the Autocasters Grass Roots poll (Crossley 1937: 24). Moreover, the response rate to the *Digest* poll in 1936 was 24 percent, up from 15 percent in 1932. Finally, if the 1936 respondents' vote recall of their 1932 vote had been used to describe the 1932 election, Roosevelt would have received 48 percent of the vote instead of the 57 percent he garnered, indicating an excess of Republican respondents. These facts suggest that Landon backers were highly motivated.

33   For a description of these issues, see Igo 2006: 118–20.

34   Age was often too sensitive to ask directly (Igo 2006: 119). Income was never asked directly; the interviewer had to ascertain whether the respondent was above, at, or below average or on relief. Moreover, there are no census data on income in those years.

35   The sample used by Gallup for his electoral polls (1936) was considerably larger than that used for issue polls like the May 1937 one analyzed in this study ($n = 2,945$, of which $n = 2,885$ represent voters). According to one source, Gallup used a sample of $n = 125,000$ for his electoral forecast of November 1936 (Robinson 1937: 48).

36   All 1936 values presented here are obtained by linear interpolation based on figures from the 1930 and 1940 censuses.

37   Willcox (1931: 245) states that men were 60 percent of the voters. He spent part of his career at the Bureau of the Census (Leonard 1961: 16).

38   See Alpern and Baum 1985; Andersen 1979: 22, 40; Kleppner 1982; and Ogburn and Goltra 1919. Most of this research deals with the 1920s, not the 1930s.

39   The results are weighted (raking) using Willcox's estimate on gender. See n. 37.

40   Based on the poll's raw results, among the voters aged 21 to 34, Roosevelt is favored by more than two to one over Landon; among the oldest, aged 55+, "only" 54 percent say that they voted for him.

41   The census age distribution is: 21–24 = 11.4 percent; 25–34 = 25.7 percent; 35–44 = 22.3 percent; 45–54 = 18.2 percent; 55+ = 22.5 percent.

42   I assume that the individuals in question were all voters. Of the original sample (n = 2,945) for the May 1937 poll, 76.8 percent are declared voters; 9.1 percent stated that they did not vote; 3.6 percent chose not to answer the question; 9.6 percent were too young to vote; and 0.8 percent were of undetermined age (16 voted for Roosevelt, 7 voted for Landon, and 1 did not vote)—these last were excluded from the analysis.

43   I assume here that voters in the 1930s and those in the 1950s through the 1980s (Wright's time series) behave similarly. Wright (1993: 310) notes that the "bandwagon effect" is more likely to occur right after the election and that it decreases with time.

44   The poll did not include a presidential performance question. Two other questions dealt with the presidency. One (question 2) asked respondents whether Congress should pass "the President's Supreme Court plan" (Roper Center for Public Opinion Research 2003a: 3): 58 percent said no (Gallup and Robinson 1938: 378). The second (question 7) asked respondents whether they would vote for Roosevelt if he were a candidate "for a third term in 1940" (Roper Center for Public Opinion Research 2003a: 3): only 35 percent said yes, 45 percent said no, and 20 percent were undecided or had no opinion. The question regarding vote choice in the 1936 presidential election was the last one asked in the survey (question 9).

45   Berinsky (2006: 516) writes rather optimistically: "There is no reason to suspect that the members of deliberately underrepresented groups—such as women and southerners—who were interviewed were systematically different from the members of those groups who were not interviewed. . . . After all, the sample imbalance exists because the pollsters deliberately drew nonrepresentative samples based on these characteristics." Berinsky misuses the word *draw*. Pollsters in those days did not "draw" samples; they set quotas and gave their interviewers instructions on how to fill them.

46   The Roper poll, carried out under the auspices of *Fortune* magazine, came within 1 point of the actual total. Unlike the other two polls (Crossley and Gallup), it was based on a national sample as opposed to an aggregate of state samples.

47   Gallup never used the data from this poll when discussing the *Digest* poll (this "oversight" is the topic of another study in progress), as he did so often, and Cahalan, who went to work for Gallup shortly after graduating from Iowa (Ohmer 2006: 82), never mentioned the results regarding the voting preference of automobile and telephone owners.

48   The company Harris Interactive, for example, claimed to have a panel of over
      6 million individuals (www.harrisinteractive.com/partner/hpolpanel.asp, accessed
      November 17, 2007; the URL is no longer available).
49   For an example of the sophistication of these tools, see Lee 2006.

References

Alpern, Sara, and Dale Baum (1985) "Female ballots: The impact of the Nineteenth
      Amendment." Journal of Interdisciplinary History 16 (1): 43–67.
American Institute of Public Opinion (1938) The New Science of Public Opinion Mea-
      surement. Princeton, NJ: American Institute of Public Opinion.
Andersen, Kristi (1979) The Creation of a Democratic Majority, 1928–1936. Chicago:
      University of Chicago Press.
Babbie, Earl (1990) Survey Research Methods, 2nd ed. Belmont, CA: Wadsworth.
Baum, Matthew A., and Samuel Kernell (2001) "Economic class and popular support
      for Franklin Roosevelt in war and peace." Public Opinion Quarterly 65 (2): 198–229.
Berinsky, Adam J. (2006) "American public opinion in the 1930s and 1940s: The analysis
      of quota-controlled sample survey data." Public Opinion Quarterly 70 (4): 499–529.
Bernstein, Stephen, and Ruth Bernstein (1999) Elements of Statistics I: Descriptive
      Statistics and Probability. New York: McGraw-Hill.
Blondiaux, Loïc (1998) La fabrique de l'opinion. Paris: Seuil.
Bradburn, Norman M., and Seymour Sudman (1988) Polls and Surveys: Understanding
      What They Tell Us. San Francisco: Jossey-Bass.
Bryson, Maurice C. (1976) "The Literary Digest poll: Making of a statistical myth."
      American Statistician 30 (4): 184–85.
——— (1977) "Reply." American Statistician 31 (4): 181.
Burner, David, and Ross Rosenfield (2003) "Polling," in Stanley I. Kutler (ed.) Dictionary
      of American History, 3rd ed., vol. 6. New York: Scribner: 409.
Burnham, Walter D. (1987) "The turnout problem," in A. J. Reichley (ed.) Elections
      American Style. Washington, DC: Brookings Institution: 97–133.
Cahalan, Don (1989) "The Digest poll rides again!" Public Opinion Quarterly 53 (1):
      129–33.
Cahalan, Don, and Norman C. Meier (1939) "The validity of mail-ballot polls." Psycho-
      logical Record 3 (1): 3–11.
Converse, Jean M. (1987) Survey Research in the United States: Roots and Emergence,
      1890–1960. Berkeley: University of California Press.
——— (2009) Survey Research in the United States: Roots and Emergence, 1890–1960.
      Repr. with new introduction. New Brunswick, NJ: Transaction.
Crossley, Archibald M. (1937) "Straw polls in 1936." Public Opinion Quarterly 1 (1):
      24–35.
Deming, W. Edwards (1964 [1943]) Statistical Adjustment of Data. New York: Dover.
Deming, W. Edwards, and Frederick F. Stephan (1940) "On a least squares adjustment of

a sampled frequency table when the expected marginal totals are known." Annals of Mathematical Statistics 11 (4): 427–44.

Desrosières, Alain (1998) The Politics of Large Numbers: A History of Statistical Reasoning. Cambridge, MA: Harvard University Press.

Erikson, Robert S. (1976) "The relationship between public opinion and state policy: A new look based on some forgotten data." American Journal of Political Science 20 (1): 25–36.

Erikson, Robert S., and Kent L. Tedin (1981) "The 1928–1936 partisan realignment: The case for the conversion hypothesis." American Political Science Review 75 (4): 951–62.

Fallacy Files (n.d.) "Fallacy watch: How to read a poll; Scientific versus self-selected," www.fallacyfiles.org/readpoll.html (accessed April 29, 2009).

Field, Mervin D. (1983) "Political opinion polling in the United States of America," in Robert M. Worcester (ed.) Political Opinion Polling: An International Review. London: Macmillan: 198–222.

Fischer, Claude S. (1992) America Calling: A Social History of the Telephone to 1940. Berkeley: University of California Press.

Gallup, George H. (1938) "Government and the sampling referendum." Journal of the American Statistical Association (201): 131–42.

——— (1972a) "Opinion polling in a democracy," in Judith M. Tanur (ed.) Statistics: A Guide to the Unknown. San Francisco: Holden-Day: 146–49.

——— (1972b) The Sophisticated Poll Watcher's Guide. Princeton, NJ: Princeton Opinion Press.

Gallup, George H., and Saul F. Rae (1940) The Pulse of Democracy: The Public-Opinion Poll and How It Works. New York: Simon and Schuster.

Gallup, George H., and Claude E. Robinson (1938) "American Institute of Public Opinion—Surveys, 1935–38." Public Opinion Quarterly 2 (3): 373–98.

Gidengil, Elisabeth (1992) "Canada votes: A quarter century of Canadian national election studies." Canadian Journal of Political Science/Revue canadienne de science politique 25 (2): 219–48.

Grier, David Alan (2002) "Predicting the unpredictable election." Chance 15 (2): 30–35.

Herbst, Susan (1993) Numbered Voices: How Opinion Polling Has Shaped American Politics. Chicago: University of Chicago Press.

Higgins, Sean (2010) "George Gallup's visionary grip on American wins and whims." Investors.com, www.investors.com/NewsAndAnalysis/Article.aspx?id=52 6858 (accessed March 11, 2010).

Hofstadter, Richard, William Miller, and Daniel Aaron (1959) The American Republic. Vol. 2, Since 1865. Englewood Cliffs, NJ: Prentice-Hall.

Huff, Darrell (1954) How to Lie with Statistics. New York: Norton.

Igo, Sarah E. (2006) "'A gold mine and a tool for democracy': George Gallup, Elmo Roper, and the business of scientific polling, 1935–1955." Journal of the History of Behavioral Sciences 42 (2): 109–34.

Kalton, Graham, and Ismael Flores-Cervantes (2003) "Weighting methods." Journal of Official Statistics 19 (2): 81–97.

Katz, Daniel, and Hadley Cantril (1937) "Public opinion polls." Sociometry 1 (1–2): 155–79.

Kircher, Steve (2007) "Pigs, pols, and polls." St. Petersburg Times, www.sptimes.com/2007/12/09/Opinion/Pigs__pols_and_polls.shtml (accessed December 9, 2007).

Kitchens, Larry J. (1987) Exploring Statistics: A Modern Introduction to Data Analysis and Inference. St. Paul, MN: West.

Kleppner, Paul (1982) "Were women to blame? Female suffrage and voter turnout." Journal of Interdisciplinary History 12 (4): 621–43.

Kruskal, William, and Frederick Mosteller (1979a) "Representative sampling, I: Nonscientific literature." International Statistical Review 47 (1): 13–24.

——— (1979b) "Representative sampling, II: Scientific literature, excluding statistics." International Statistical Review 47 (2): 111–27.

——— (1979c) "Representative sampling, III: The current statistical literature." International Statistical Review 47 (3): 169–95.

——— (1980) "Representative sampling, IV: The history of the concept in statistics, 1895–1939." International Statistical Review 48 (2): 245–65.

Ladd, Everett Carll (1985) The American Polity: The People and Their Government. New York: Norton.

Ladd, Everett Carll Jr., with Charles D. Hadley (1975) Transformations of the American Party System. New York: Norton.

Lee, Sunghee (2006) "Propensity score adjustment as a weighting scheme for volunteer panel web surveys." Journal of Official Statistics 22 (2): 329–49.

Leonard, William R. (1961) "Walter F. Willcox: Statist." American Statistician 15 (1): 16–19.

Literary Digest (1928) November 3.

——— (1936a) August 22.

——— (1936b) August 29.

——— (1936c) October 31.

——— (1936d) November 14.

Moon, Nick (1999) Opinion Polls: History, Theory, and Practice. Manchester: Manchester University Press.

New York Times (1936) "Digest to seek reason for failure of poll." November 4.

News-Week (1936) "Poll: Dr. Gallup closes a gap between people and government." November 14.

Oberschall, Anthony (2008) "The historical roots of public opinion research," in Wolfgang Donsbach and Michael W. Traugott (eds.) The Sage Handbook of Public Opinion Research. Thousand Oaks, CA: Sage: 83–92.

Ogburn, William F., and Inez Goltra (1919) "How women vote." Political Science Quarterly 34 (3): 413–33.

Ohmer, Susan (2006) Gallup in Hollywood. New York: Columbia University Press.

Ozouf, Jacques (1963) "L'enquête d'opinion en histoire. Un exemple: L'instituteur français, 1900–1914." Le mouvement social 44: 3–22.

Raj, Des (1972) The Design of Sample Surveys. New York: McGraw-Hill.

Robinson, Claude E. (1932) Straw Votes: A Study of Political Prediction. New York: Columbia University Press.

——— (1937) "Recent developments in the straw-poll field." Public Opinion Quarterly 1 (1): 45–56.

——— (1939) "Measuring public opinion." Savings Bank Journal 20 (4): 2–4.

Robinson, Daniel J. (1999) The Measure of Democracy: Polling, Market Research, and Public Life, 1930–1945. Toronto: University of Toronto Press.

Roper, Elmo (1940) "Sampling public opinion." Journal of the American Statistical Association (210, pt. 1): 325–34.

Roper Center for Public Opinion Research (2003a) Gallup Poll no. 83.

——— (2003b) USAIPO1937-0083.

Rossi, Peter H., James D. Wright, and Andy B. Anderson, eds. (1983) Handbook of Survey Research. New York: Academic.

Rugg, Donald (1947) "How representative are 'representative samples'?," in Hadley Cantril (ed.) Gauging Public Opinion. Princeton, NJ: Princeton University Press: 143–49.

Rusciano, Frank Louis (2007) "Book Reviews: Susan Ohmer; George Gallup in Hollywood." Public Opinion Quarterly 71 (2): 314–18.

Saxon, Wolfgang (2000) "Seymour Sudman, 71, expert in survey design." New York Times, May 8.

Shively, W. Phillips (1972) "A reinterpretation of the New Deal realignment." Public Opinion Quarterly 35 (4): 621–24.

Smith, Charles W., Jr. (1952) "Measurement of voter attitude." Annals of the American Academy of Political and Social Science (283): 148–55.

Squire, Peverill (1988) "Why the 1936 *Literary Digest* poll failed." Public Opinion Quarterly 52 (1): 125–33.

Stephan, Frederick F. (1939) "Representative sampling in large-scale surveys." Journal of the American Statistical Association (206): 343–52.

——— (1942) "An iterative method of adjusting sample frequency tables when expected marginal totals are known." Annals of Mathematical Statistics 13 (2): 166–78.

US Bureau of the Census (1931) Statistical Abstract of the United States: 1931. Washington, DC: US Government Printing Office.

——— (1943) Statistical Abstract of the United States: 1942. Washington, DC: US Government Printing Office.

——— (1970) Statistical Abstract of the United States: 1970. Washington, DC: US Government Printing Office.

——— (1997) Historical Statistics of the United States on CD-ROM: From Colonial Times to 1970, ed. Susan B. Carter. New York: Cambridge University Press.

Weiner, Sanford L. (1976–77) "The competition for certainty: The polls and the press in Britain." Political Science Quarterly 91 (4): 673–96.

Weisberg, Herbert F., Jon A. Krosnick, and Bruce D. Bowen (1998) An Introduction to Survey Research, Polling, and Data Analysis, 3rd ed. Thousand Oaks, CA: Sage.

Wilks, Samuel S. (1940) "Representative sampling and poll reliability." Public Opinion Quarterly 4 (2): 261–69.

Willcox, Walter F. (1931) "An attempt to measure public opinion about repealing the Eighteenth Amendment." Journal of the American Statistical Association, n.s. (175): 243–61.

Wright, Gerald C. (1993) "Errors in measuring vote choice in the national election studies, 1952–88." American Journal of Political Science 37 (1): 291–316.

Zeisel, Hans (1985) Say It with Figures, 6th ed. New York: Harper and Row.

Zunz, Oliver (1998) Why the American Century? Chicago: University of Chicago Press.