

From Alchemy to Home-Brewed Chemistry—Polling Transformed

by Bernard Roshco and Irving Crespi

On a Sunday in the fall of 1935, the third year of Franklin D. Roosevelt's first term as president, a blimp cruised over Washington, DC, towing a streamer that proclaimed, *America Speaks!* The mobile billboard was promoting a newspaper column making its debut in the *Washington Post*. The column was also being introduced—without airborne announcements—in 39 other newspapers. *Newsweek* magazine called *America Speaks!* “probably the most ambitious newspaper feature ever devised.” A new, nationwide poll would ask Americans their views on then-current issues and report the findings weekly under the byline of a 34-year-old advertising-agency executive whose name would eventually be so familiar that it became a synonym for public-opinion polls, George H. Gallup.

The times were propitious for polling, given the crises and controversies that racked the country. The stock market crash of 1929 and the subsequent Great Depression made Republican Herbert Hoover a one-term president. In November 1932, Franklin D. Roosevelt became the first Democratic Party candidate to gain the presidency since Woodrow Wilson's election in 1916. By the time FDR took office in 1933, one in four previously employed workers was out of a job and 40% of the commercial banks had folded. In its “first hundred days,” FDR's New Deal administration enacted a spate of emergency legislation and created a slew of agencies. The Federal government hired the unemployed, subsidized farmers, closed and then reopened the banks. The White House and the Supreme Court were soon at loggerheads as the Court struck down various New Deal measures as unconstitutional. There was plenty of scope for a pollster's questions.

The *Literary Digest's* Straw Polls

To give his findings more authority, Gallup presented them as polls conducted by the American Institute of Public Opinion. But to most newspaper editors, politicians, and the general public the premier name in polling was the *Literary Digest*. A weekly whose founding in 1890 made it the granddaddy of later news magazines, the *Digest* had conducted polls

“*Gallup concluded that the Digest's sample, biased toward higher-income, Republican respondents, would make a major mistake in predicting the outcome of the 1936 Presidential election.*”

since 1916, when it discovered they were an effective way to solicit subscriptions. From then on, the *Digest* did mail-order polling on an ever-larger scale, sending millions of questionnaires, combined with subscription forms, to potential readers across the country.

In its first presidential poll, in 1924, the *Digest* mailed more than sixteen million “ballots”; by 1932, the mailing was up to twenty million.

Various newspapers had been doing such “straw polls” since the mid-nineteenth century, but on a much smaller scale. From about 1900 on, they were an increasingly frequent newspaper feature. Mostly, they asked a single question on a local political issue. The polling was usually confined to a single city, occasionally an entire state.

Most straw polling was haphazard and hasty. Local newspaper reporters were often the poll-takers. To collect as many “votes” as quickly as possible, “ballots” were passed out where crowds gathered and then collected in the “ballot boxes.” Some newspapers conducted straw polls by printing coupons in the paper and inviting readers to mail them back, paying no attention to who actually answered the questions or how often they answered them.

Establishing a “Scientific” Method

Whether ballots were accumulated in person or by mail, straw-polling was catch-as-catch-can, done mostly to publicize the sponsoring publications. But, beginning in the 1920s, publishers and other businessmen with commercial reasons for wanting to know their customers' preferences turned to market research. Market researchers paid careful attention to the cross-section of the public they were sampling, designating “quotas” of respondents according to age, gender, income, and other categories deemed relevant. Quota sampling sought to match the sample being polled with the population of customers whose buying habits and product choices were being studied.

Though unknown to the general public, Gallup was respected among market researchers for his “Gallup method” of learning what interested newspaper and magazine readers. Gallup developed and tested his method at the University of Iowa, where he and his assistants interviewed 1,000 readers of the *Des Moines Register and Tribune* to learn which of the newspaper's features were most popular. He divided his sample of 1,000 readers into five income categories, further divided by sex, and by urban or rural residence. Gallup's innovation was to have the interviewers hand readers a newspaper and ask what they actually read. To the chagrin of editors, who relied on their own intuitions about reader preferences and were ready to discard most of their comic strips, Gallup found that comics and photographs attracted many more readers than front-page stories.

Gallup refined his approach to sampling and questioning with readership surveys for various publications. By 1932, his growing reputation in market research landed him a job as an executive with Young and Rubicam, a leading advertising agency. Privately, he tested applications of market-research techniques to politics. If he could determine reader preferences, why not voter preferences?

Up to then, only the *Literary Digest* did nationwide, pre-election polling to any meaningful extent. Before a Presidential election, the *Digest* carried out a single, massive poll. Ballots were mailed several months before the election, and cumulative returns were released periodically, as more ballots came back. A final prediction, based on all the returned ballots, was published just before Election Day. Each such poll asked respondents which candidate they favored and who they had voted for in the previous election. Because the *Digest* tried to call the Electoral College vote by predicting the outcome in every state, the questionnaires asked respondents in which state they lived.

Gallup intended to compete directly against the *Digest's* poll by calling the Electoral College vote in every state. Since he claimed he could sample national opinion every ten days, his problem was to be more accurate than the *Digest* with much smaller samples. Experimenting with nationwide polls since 1933, Gallup used the Congressional elections of 1934 as a key test of his sampling methods. For the first time in the twentieth century, an incumbent administration wound up with a larger Congressional majority in an off-year election. Reviewing the demographic breakdowns from his unpublished 1934 polls, Gallup concluded that the *Digest's* sample, biased toward higher-income, Republican respondents, would make a major mistake in predicting the outcome of the 1936 Presidential election.

America Speaks! or Your Money Back

Gallup offered his new poll to potential newspaper clients with a guarantee that his forecast of the 1936 Presidential election would be closer to the final count than the *Digest* poll, or he would refund the entire cost of running *America Speaks!* The offer seemed reckless, given the *Digest's* near-perfect 1932 prediction, when it called Roosevelt's margin of victory over Hoover within less than one percentage point of the actual popular vote.

“***The name Gallup gave to his column, America Speaks! signified his interest in auditing and reporting the voice of the people. For him, predicting elections was the most effective way to provide evidence that his polls expressed what the public was thinking.***”

The first release of *America Speaks!* was published in forty papers on October 20, 1935. After two-and-a-half years of New Deal programs, Gallup asked, “Do you think expenditures by the Government for relief (i.e., welfare) and recovery are too little, too great, or just about right?” A majority of the public, 60%, thought the government was spending too much. Only 9% thought it was spending “too little” and 31% thought it was “about right.”

The Roosevelt Administration's chief press spokesman, Charles Michelson, loftily dismissed the upstart pollster and his data, declaring: “You will look in vain for the name of Dr. George Gallup in *Who's Who*.” Michelson dismissed the Gallup finding as a Republican scheme to discredit the Administration and advised readers to await results from the country's best-known poll, the one conducted by the *Literary Digest*. It was, Michelson de-

clared, “honest, unbiased and unmanipulated.”

Undeterred, Gallup continued to ask questions on news-making issues. Unlike the straw-pollers, he “pre-tested” questions by sending out interviewers to pose alternatively worded queries in order to learn which versions were more understandable and germane to popular concerns. During the following months, findings presented in *America Speaks!* suggested that Americans had mixed feelings about the New Deal. Even if they opposed government spending in general, they favored particular programs. In November, Gallup asked, “Should the Federal government pay the veterans their bonus now?” Fifty-five percent said “yes.” In December, when Gallup asked, “Are you in favor of Government old-age pensions for needy persons,” 89% said “yes.” But, a month later, when he asked, “Do you think it is necessary at this time to balance the budget and start reducing the national debt,” 70% also said “yes.” When those saying “yes” were asked if it should “be done by government economies, higher taxes, or both,” 80% opted for government economies, 2% for higher taxes, and 18% for both.

Asking questions on current public issues at frequent intervals and presenting the responses as news marked a significant innovation in polling. Ultimately, it would have a major impact on politics, as was foreshadowed by Michelson's instant counterattack on Gallup's first column. The name Gallup gave to his column, *America Speaks!* signified his interest in auditing and reporting the voice of the people. For him, predicting elections was the most effective way to provide evidence that his polls expressed what the public was thinking.

Predicting the *Digest's* Results

By late June of 1936, the presidential campaign was underway. FDR was running for a second term. To run against him, the Republican Party nominated Governor Alfred M. Landon of Kansas.

Polling America

Shortly after the Republican convention, Gallup reported the responses to this question: "If there were only two political parties in this country—Conservative and Liberal—which would you join?" "Conservative," said 53%; "Liberal," said 47%. Whatever these answers seemed to imply about the coming election, Gallup had been reporting majority support for FDR.

Within a month of the Roosevelt and Landon nominations, Gallup topped his promise that his unproven poll would forecast the coming election more accurately than the *Digest*. He predicted the outcome of the *Digest* poll in his column of July 12, 1936, weeks before it began mailing out its millions of ballots: "If the *Literary Digest* were conducting its poll at the present time, following its usual procedure, Landon would be shown in the lead. The actual figures would be in the neighborhood of 44% for Roosevelt, 56% for Landon." The *Digest* began mailing ten million ballots in August. (Facing financial stringencies, it halved its 1932 mailing.) With a sample only a small fraction of the size the *Digest* compiled, Gallup came within one percentage point of the final results the *Digest* would announce three-and-a-half months later.

Gallup then explained how he arrived at his figures: "Since the Institute of Public Opinion sends part of its ballots to the same lists covered by the *Literary Digest*, it is possible to predict with a high degree of accuracy the sentiment which the *Digest* will find. The lists comprise telephone subscribers, automobile owners, and registered voters. The upper economic levels are represented to a much greater extent than the lower levels in such lists, and particularly in the returned ballots. People at the upper levels are more inclined to answer ballots than people at the lower end of the economic scale. When the lower one-third of the voting population is fully represented, Roosevelt's percentage changes from 44% to 52%."

Gallup had recognized that the *Digest* sample was doubly flawed. Not only was it unrepresentative of the American public, but those who responded were not an accurate cross-section of the sample. Those most opposed to New Deal measures and most likely to support Landon sent back their *Digest* ballots in disproportionate numbers to their actual representation in the sample. If everybody in the sample had actually responded, the result would have been substantially different.

The *Digest* was not concerned with statistical niceties. It held to the popular belief that the largest possible sampling would be the most representative, hence the most accurate. Gallup recognized how to take advantage of the *Digest*'s sampling method. Since he was predicting the outcome of the *Digest*'s poll, not the outcome of the actual election, he did not have to poll a sample of the entire public; he had to poll only the *Digest*'s sample. By using the same lists as the *Digest*, he could duplicate its poll.

Although Gallup also drew a large part of his sample from lists, about one-third of his polling was done in person. He sent interviewers out to button-hole people on relief and others not likely to show up on mailing lists of the relatively affluent. In contrast to the *Digest*'s broadside approach, Gallup set up a staff that compiled a population profile for each state, in order to determine the sampling quotas. Returns were tabulated for each state according to the quota categories, so that the responses of various population groups could be reported.

Return Shots Fired

Charlie Michelson didn't complain about Gallup's July 12 column. But, the *Literary Digest* did. Its outraged editor, Wilfred J. Funk, distributed a letter to the press, advising "our fine statistical friend" that the *Digest* would continue to employ "those old-fashioned methods that have produced correct forecasts exactly one hundred percent of the time." Observers not involved in actual polling also criticized the *Digest* for its skewed sample. The Washington columnist of the *New York Times*, Arthur Krock, wrote in early September that "two of the

“ Separately, the trio of Gallup, Crossley and Roper transformed polling from alchemy to home-brewed chemistry. They constantly revised their samples, seeking to make them more representative of the voting-age population. ”

four groups which will probably decide the election have hardly been sampled, if at all...the new voters, the colored voters, labor and agrarian citizens; and it is doubtful whether it will be possible for the poll-takers to produce an accurate representation of the first two... because their names are not so easily come by in the public records."

Funk responded immediately with a letter to the *Times*: "As you doubtless know, our polls are secret and we have no knowledge whatsoever about the economic status of any of the poll voters as they vote. We know, however, that voters in the four groups Mr. Krock enumerates have cast ballots in the *Literary Digest* poll because they have taken the trouble to write us separate communications expressing their views."

Funk finished his letter with a sneer at a "so-called 'scientific' poll which is syndicated among a number of newspapers. What did this 'scientific' poll forecast in 1932? In 1928? In 1924? Did it forecast the break-up of the solid South in 1928, when seasoned political observers believed such an upset to be almost impossible? The answer, of course, is negative because the 'scientific' poll is itself unseasoned, has never had to forecast against a national election. May we

add that in 1932 the *Literary Digest* forecast a major party vote ratio of 59.86 per cent for Mr. Roosevelt, whereas the official election showed he obtained 59.14 per cent of the Roosevelt-Hoover vote, indicating a deviation of 0.72 per cent.”

Two Upstarts Named Crossley and Roper

The *Digest* poll’s credibility with the public and most politicians during the 1936 campaign was based on its remarkable record in the 1932 election. In the face of that achievement, the statistically unsophisticated were not prepared to be critical. Although Gallup was the sole target of Funk’s attack, he was not the only proponent of a radically different approach to sampling. Two other market researchers had also taken up political polling, Archibald Crossley and Elmo Roper. Well-known in market research, their commercial work brought them invitations to try political polling.

Archibald Crossley had worked in the research department of the *Literary Digest* between 1922 and 1926, but was not involved with its poll. He then founded his own firm and became noted for radio-audience research. From regular telephone interviews with radio listeners in fifty cities, he developed reports of audience-size, for hour and day, and the composition of audiences according to their purchasing power, sex, and age. By 1931, he was measuring the popularity of programs and audiences in 150 cities. So by 1936, when he was asked to do pre-election polling for the Hearst papers, he was amply prepared to survey voters.

Elmo Roper went into market research after failing in the retail jewelry business in the 1920s. Like Gallup, who had been skeptical of editors’ intuitive certainties about reader preferences, Roper was dubious about the market certainties of retail jewelers. He switched his business to wholesaling stock to other jewelers and interviewed them and their customers as he criss-crossed the coun-

try. After recounting his findings to a friend in advertising, he was introduced to a Harvard professor of marketing. The outcome was a new market-research firm and a meeting with Henry Luce, publisher of *Time*, *Life*, and *Fortune*. This led to the *Fortune* Poll, which made its debut in the spring of 1935, a few months before *America Speaks!*

Unlike the weekly reports by Crossley and Gallup, Roper’s pre-election polls for *Fortune* were published quarterly. Each of the three had his own formula for determining the right number of respondents for various demographic categories. Unlike Gallup, Roper and Crossley depended entirely on personal interviews. Both were in complete accord with Gallup that the *Digest*’s approach was flagrantly flawed, although they were far less imaginative than Gallup in publicizing the fact.

Separately, the trio transformed polling from alchemy to home-brewed chemistry. They constantly revised their samples, seeking to make them more representative of the voting-age population. There was not much “science” in the sense of statistical theory in quota sampling, but a great deal of meticulous, laborious effort went into designing samples and tabulating responses. Compared to the happenstance approach of straw-polling, which allowed respondents to return multiple ballots, adapting pre-designated quotas to political polling was revolutionary.

All three employed far smaller samples than the multi-state straw polls. Roper, who was not attempting state-by-state predictions, used the smallest samples, 4,500 per poll. Crossley polled at two-week intervals from August through October, each time with a sample of 30,000. Gallup, the only one of the three who used mailed ballots as well as personal interviews, gathered the largest samples in order to compete with the *Digest*’s state-by-state forecasts. In addition to personal interviews, he mailed out at least 100,000 ballots each time. Nonetheless, the total samples of all three for the entire campaign were a

fraction of the *Literary Digest*’s mega-mailing.

More Straw Polls

1936 was a banner year for election predictions. *Time* wrote that “never before in US history have so many extensive and intensive attempts been made so far in advance to foretell what will happen on Nov. 3.” The *New York Times* commented, “Polling is obviously headed for the role of this nation’s major seasonal industry.”

With both parties attempting to roll bandwagons, trivial results were offered as portents of the election. Three members of the Democratic State Committee from Landon’s native Kansas polled dining-car waiters while they traveled to a Democratic Party conference. They reported 14-0 for Roosevelt. A straw poll of Yale students found they favored Landon, 1,183 to 550.

The *Farm Journal*, which like the *Literary Digest* had been conducting straw polls since 1916, queried 164,000 farmers in 34 states and found Landon favored in 21 of them. The magazine noted that its straw polls had correctly forecast the winner of every election since 1916. But, it warned, “these figures include practically no town or city ballots, and it is certain that in many States the urban vote will be cast in opposition to that of the country.” The Grass Roots Poll was conducted by a consortium of 3,000 small-town newspapers in 39 states. The papers published ballots that had to be cut out and mailed. A return of 919,441 ballots showed a 61% majority for Landon.

The *Baltimore Sun* conducted the ultimate state-wide straw poll. It attempted a virtual census by mailing a postcard ballot to every registered voter in Maryland. More than three-quarters of a million voters were registered in Maryland, with a two-to-one preponderance of Democrats. Less than one-third of those polled by the *Sun* returned ballots. But, since the sample comprised the entire electorate and, there-

Polling America

fore, was preponderantly Democratic, a sizable pro-Roosevelt margin was recorded. The *Sun*'s poll showed 64% for Roosevelt and 36% for Landon. In November, Roosevelt received just under 63% of the actual Maryland vote.

On the other hand, the *Literary Digest* polled a virtual neck-and-neck race in Maryland. Reporting 51% for Roosevelt and 49% for Landon, the *Digest* barely avoided a mis-call. In states where the margin was closer, the *Digest* was misled by the biases of its sample. Gallup revealed the extent to which income influenced candidate preference when he published the results of his poll of August 23, 1936; those who owned telephones were 59% for Landon, owners of autos were 56% for Landon, those on relief were 18% for Landon. On September 4, the *Digest* released its earliest returns, which showed Landon with a substantial lead. But, the *Digest* warned, the result was too preliminary to be conclusive. As more returns came in, they were added to the running total, which continued to show Landon in the lead.

Meanwhile, Gallup, Crossley, and Roper released a succession of individual polls, each conducted with a separate sample. All showed Roosevelt in the lead. These successive polls were a major innovation. Because they interviewed a new sample each time, the trio could track whether opinion changed over time, in the wake of significant events. Gallup, polling after Landon's nomination and again after Roosevelt's acceptance speech to the Democratic convention, noted shifts in popular support after the two events, as well as increasing Roosevelt strength as election day approached. The *Digest*'s cumulative returns could not track trends; though its ballots had been mailed out over a period of weeks and came back over a period of months, all were part of a single sample.

Both Gallup and Crossley found themselves confronting unhappy publishers. One of Gallup's major clients, the *New York Herald-Tribune*, arch-rival of the *New York Times* and pillar of the Republican Party, banished Gallup's reports to deep inside the paper and juxtaposed them with straw polls from around the country that showed Landon ahead. Other client papers also kept Gallup off the front page. The Hearst papers, Crossley's sole clients, took their editorial cue from publisher William Randolph Hearst, who had become a fierce Roosevelt foe. Crossley was likewise banished to the inside pages.

The Outcome

The *Literary Digest*'s final forecast was based on 2,376,523 returned ballots, less than one-quarter of those mailed. On October 29, the *Digest* predicted Landon would carry 32 states and Roosevelt 16, giving Landon 370 electoral votes and Roosevelt 161. It predicted the popular vote would divide approximately 57% for Landon and 43% for Roosevelt. Landon actually got about 38% and Roosevelt a little less than 61%. Landon carried two states, Maine and Vermont, giving him eight electoral votes to Roosevelt's 523.

All three "scientific" pollsters called the winner correctly. But their estimates of Roosevelt's popular vote diverged. Gallup and Crossley under-estimated his margin by about 7%. Roper came closest, over-estimating his margin by about 1%. All three recognized that their methods needed substantial revision and improvement, despite the triumph over straw-polling. Speaking to a marketing convention several weeks after the election, Crossley highlighted the difficulty of determining sampling quotas for each population group. The *Literary Digest* simply declared itself bewildered: "Do Republicans live nearer mail-boxes? Do Democrats generally disapprove of straw polls? We don't know that answer."

The *Digest* did not survive to seek an answer. Before the next Presidential election, the magazine was defunct. The problem of determining appropriate quotas was eventually "solved" by largely discarding quota sampling. In 1948, after the polling community unanimously mis-called the Presidential election by predicting a victory for Dewey over Truman, polling became more "scientific" by examining its mistakes. Polling organizations adopted new sampling techniques, based on more rigorous statistical methods than filling estimated quotas. Since polling's showcase effort was election forecasts, methods were developed to estimate each respondent's likelihood of actually voting. To stay abreast of last-minute changes in voting intentions, polls were taken up to the eve of an election.

From 1936 to 1996

Sixty years later, journalists, politicians, and social scientists have made poll findings an integral part of news, politics, and political analysis. Nowadays, no political professional would trust the findings of a poll that resembled the *Literary Digest*'s efforts. At the same time, nationwide straw-polling has revived and become a money-maker. Telephones and computers make it easy and economical to conduct straw polls that cover the entire United States. Instead of the hundreds of clerks that the *Literary Digest* had to employ to mail and tabulate ballots, various television shows now poll by using "900" telephone numbers: Do you think the president should do "X" or should he do "Y?" If you favor "X," you are invited to dial a "900" number. If you favor "Y," dial a different "900" number. A computer tallies the "votes," and dialing incurs a charge. You are part of the "sample" as many times as you're willing to pay. The "scientific" polling fostered by Gallup, Crossley and Roper transformed politics; the current successors to the *Literary Digest* poll transform straws into cash.



Bernard Roshco is past director of the Office of Opinion Analysis, US Department of State and Irving Crespi is past executive vice president, The Gallup Organization