

Some Things That Concern Me

BURNS W. ROPER

IT HAS BEEN said that opinion research is part art and part science. I agree. I have heard it said that opinion research is *half* art and *half* science. I would say that a good deal more than half is art and correspondingly less than half is science. Or maybe it would be more accurate to say that substantially more than half of what has been *neglected* in opinion research is on the art side and most of what has been *addressed* has been on the science side.

People feel comfortable with the science side. It is "hard" rather than "soft." It is "measurable." "With a sample of this size, the results of this survey have a sampling error of plus or minus 2.7 percentage points at the 95 percent confidence level." That is a precise, solid, factual statement. (And it suggests that the person making that statement is erudite.) Never mind that we often don't know whether this is one of the 95 surveys that is within this margin, or one of the five that is outside it. Or that most laymen—highly intelligent laymen—are misled into thinking that such an error statement is a measure of *total* error, not just sampling error.

Such a statement is factual and scientific, and therefore nonarguable. As such, it is a comforting, cuddle blanket. (A correlation of 0.87 is also good, precise, and therefore comforting.) You can't say that the odds are 95 out of 100 that a given opinion survey design will produce the right answer to a problem within plus or minus 2.7 percentage points. That becomes a matter of judgment, hence is both nonscientific and arguable.

Abstract To improve the craft of opinion research the author makes these recommendations: (1) emphasize the "art" side of opinion research, (2) avoid establishing strict standards of performance, (3) push for more disclosure than the present code calls for, and (4) substitute question wording for sampling error statement in public release of poll results.

Burns W. Roper, Chairman of The Roper Organization, Inc., was President of AAPOR in 1982–83. This is an adapted version of his presidential remark at AAPOR'S 38th Annual Conference, Buck Hill Falls, Pennsylvania, May 19–22, 1983.

If one has written a question or a set of questions to find out what people think should be done about the social security system, what mathematical probability is there that they truly tap the public's real feelings on the subject? What is the mathematical correlation between the answers to these questions and reality?

One night, a few years ago, I heard on the NBC Network Evening News that a just completed NBC poll showed 68 percent of the American public favoring the SALT II agreement. I was more than a little surprised, for that very afternoon we had gotten the printout to our latest Roper Reports. It showed 33 percent—just under half as many—favoring passage of SALT II. Along with the 68 percent result, the technically accurate and confidence-inspiring plus or minus 3-percentage-point-error statement was cited by NBC. So clearly, not less than 65 percent nor more than 71 percent favored SALT II. Yet we found only 33 percent favoring SALT II.

Subsequent checking indicated that the two interviewing periods were almost identical. Both samples were designed to be representative of the adult population of the nation, age 18 and up. Yet 68 percent approval in the NBC News poll; 33 percent approval in ours.

In due course, I received an NBC release containing the wording of their "SALT" question. It read, "Do you favor or oppose a new agreement between the United States and Russia which would limit nuclear weapons?" Note that it said "*a* new agreement" not "*the proposed* new agreement." Note too that the word "SALT" never appeared in the question. By contrast, our question told respondents some of the basic provisions of what was identified as "a SALT treaty." It pointed out that there was a lot of "controversy about this proposed treaty" and it then asked, "Do you think the U.S. Senate should vote for this new SALT treaty or against it?" Now it's a bit clearer why the two surveys got a 35-point difference in results. And clearly the answer lies in the art part, not the science part.

Or take our national vs. private health insurance question. Sometimes we ask it at the end of a series of questions about whether you have a family doctor, how well satisfied you are with your health care service in terms of quality, in terms of availability, in terms of cost, and in terms of the provisions you have for paying those costs. Sometimes we ask the question alone—by itself. When we have asked it at the end of the series on health care, private health insurance has always been favored over national health insurance, by a moderate margin. When we have asked it alone, national health insurance has always won, by an equally moderate margin. It is the art part that causes the difference, not the science part.

Eighty percent of the error in surveys lies on the art side; yet 80

percent of the articles, teachings, and caveats are on the science side. Incidentally, both of those 80 percent figures are plus or minus 6.1 percent at the one sigma level—so you can be reasonably confident about their accuracy. Fortunately, there has been some resurgence of interest in the “art” side, both in academe and in the commercial world.

Another subject I’d like to address is the matter of standards in opinion research. I don’t mean ethical standards, or disclosure standards. I mean *performance* standards—what is the “right” way, what is the wrong way; what is “acceptable,” what is not. I suppose I might be in favor of such standards if all other practitioners would come up to *my* standards of performance, but at the same time not go way beyond my standards to a ridiculous and absurd level that time, cost, and accuracy cannot justify.

My argument against performance standards is, again, that it’s a matter of judgment—art—as to what degree of rigor is called for. Example: We did a real “quick and dirty” in 1962 for John Chafee, Republican candidate for governor of Rhode Island. A mutual friend of his and my father’s called my father to see if he thought we could do anything in the remaining three weeks of the campaign that might help Chafee turn a doubtful-seeming cause into victory—and do it for \$1,500. Specifically, he wanted to know if we could find out what issues Chafee should concentrate on in the final 10 days or so of the campaign.

We sent two people up to Rhode Island and told them each to “work the crowds” in various shopping malls. (Then such a procedure had no name; now it is made to sound precise and scientific with the elegant term “intercept interviews.”) We told these two people to stop adults of all voting ages and both sexes and ask them which of the two candidates for governor they voted for in 1960, and which of the two present candidates they planned to vote for next month. If a respondent gave no indication of switching, the interviewer thanked him or her and moved on to the next passerby. But if a person indicated a likelihood of switching—Republican then, undecided now, or Democratic then, Republican now—then the interviewer was to probe for the reasons. In part, this design was based on the fact that the Democrat was the same person in both years. Why were you for him then and uncertain now? Or why were you against him then and for him now?

We got a lot of good stuff and Chafee concentrated effectively on a couple of indicated issues in the final days of the campaign. But, in addition, I was struck by an imprecise statistic. Something like 90 percent of the potential switchers were potential switchers from the

Democrat to the Republican; only about 10 percent were the other way. I looked at the 1960 election results. Having an approximate idea of how many people we screened to get the potential switchers that we found, and assuming the total group we contacted was not wholly unrepresentative of the state, we thought it looked like enough were switching, and switching *to* the Republican, so that Chafee could actually win—something few, if any, politicians in the state thought possible. Encouraged by his suddenly brightened prospects, and armed with a couple of issues important to prospective switchers, Chafee went on to win—by the slim margin of 398 votes after the absentee ballots were counted.

Nothing about that survey would have met any standards AAPOR might ever have agreed to as acceptable performance. The sampling was not probability; it was not even good quota. It was random—in the haphazard sense of the word. Everything else about the study was equally imprecise. There are few, if any, other studies that I have been responsible for in my nearly 37 years in the field that have been so unscientific. At the same time, there are few studies I can think of that have better served the purposes for which they were intended. The study provided proper hope and reassurance to the candidate, it provided him with knowledge as to what issues to stress in order to influence the switchables, it provided all this in a time frame that let him act on the results, and it did it all for the budget that was available.

I am convinced that the poll was good for at least 500 of the 164,000 votes Chafee got—or more than the 398 vote margin by which he won. He was a good governor for three terms. He later became a distinguished Secretary of Navy, and is now Rhode Island's junior senator. He is the kind of Republican I would like to see more of. Both because of the study itself, and because of the important, even if limited, role it played in Senator Chafee's career, it is one of the polls of which I am proudest.

Suppose matters reached the point where we had standards as to question wording: "Thou shalt not measure presidential approval except with the Gallup question." The Gallup presidential job performance question is a good question and one I have made considerable use of over the years—enough use, in fact, to have discovered that it doesn't measure only job performance. If it did, then President Kennedy wouldn't have gotten a higher "job performance" rating after the Bay of Pigs debacle than he did after his Cuban missile triumph.

Knowing that the Gallup question measures a mixture of job performance and emotional support, we decided to write a different presidential rating question—not necessarily better, but different—a

question that focuses more on emotional support than job performance. By asking our own question instead of the Gallup “standard” question, we learned something. All through Carter’s term—except for immediately after Iran took our embassy staff as hostages—we showed that *support* for Carter, on the basis of our question, was 10 to 12 points higher than *approval* of Carter, on the basis of the Gallup question. In other words, the person was more popular than the job he was doing. For the first two years of the Reagan administration, support for Reagan on the basis of our question has run only about one point above approval of Reagan on the basis of Gallup’s. And support on the basis of ours has run considerably behind the support for Carter that we found at comparable points in his term.

This says, then, that Carter as a person was more approved than was the job he did. Reagan, by contrast, is little more liked as a person than is the job he has done. This is the exact opposite of the frequently stated journalistic observation that Reagan is personally more popular than his program is. If standards said, “Thou shalt use the Gallup question,” this insight would never have been uncovered—at least not in this way.

Another—and rather basic—problem that concerns me is: What are and what are not appropriate areas or subjects on which to conduct opinion research? In one sense I suppose the answer is: Any subject that respondents are willing to talk about and that someone is willing to pay for to have researched. But that’s a little like saying heroin is proper commerce because someone is willing to sell it and others are willing to buy it.

Let me illustrate my concern. Earlier this year we asked people whether Congress was right to withhold production funds from the MX missile until a better basing mode than Dense Pack could be found, or whether Reagan was right in seeking production funds, or whether the MX should be scrapped entirely. Nearly half said Congress was right; only one in nine agreed with Reagan; nearly twice that many said it should be scrapped entirely. The question was balanced and fair. Moreover, I personally like the results it produced. Why, then, am I concerned?

I’m concerned because we also asked people how much they felt they knew about the MX. Only 7 percent said “a lot.” Less than two in five said either “a lot” or “only something”—as opposed to “very little.” This self-professed ignorance was confirmed by the fact that one in four had no opinion about the MX, even after we described to them Reagan’s position and the Congress’s position. Should we be measuring and reporting public opinion on a subject that so few people feel informed about? Should public opinion like this—if it can

be called public opinion—play a role in policy formation on a subject like this?

I think it really doesn't matter very much whether public opinion of this character *should* play a role, for the fact is it *will* play a role—in terms of Congress's mail, if not by reason of a poll. Maybe it's better to reveal that only 7 percent feel they know "a lot" and that 23 percent have no opinion on the issue than it is to let Congress infer the state of public opinion from its mail. But it bothers me. It also bothers me when we feel we have to write a 50- or 100-word descriptive preamble before we can make a 20-word question meaningful to respondents. I've toyed with a rule of thumb: If it takes more words to describe the situation than it does to pose the question, then the question should not be asked. Or it should be asked only of those who know enough so that they don't need the description.

But that bothers me too. It represents a kind of elitism that I don't like in any society and that seems to me inconsistent with a democracy. Moreover, who is to decide what is the proper information threshold before public opinion should be listened to and considered?

My real concern, I think, is not over *asking* questions in such areas; it is over the way poll results in such areas are both *reported* and *interpreted*. Several years ago the CBS poll reported results that were diametrically opposite to ours. I think the subject may have been gasoline rationing, and I think they found something like two-thirds of the public in favor of rationing when we found two-thirds opposed. Or maybe it was the other way around. It doesn't matter, in terms of the point I want to make. The point is we both investigated the same subject at the same time and got opposite results. CBS reported their results on air, with the usual confidence-building plus or minus 3-percentage-point-error statement. So now we know—know within 3 points, that is—what public opinion is on that important subject, don't we? No, we don't.

The *two* polls *taken together*, do tell us a good deal about public opinion on that important subject. They tell us that public-opinion is by no means decided. It is either in its formative stages and different arguments elicit different reactions, or else it's not a subject people really care about, hence, different question wordings can wildly alter the results obtained.

Measuring opinion when it is uninformed or fluid is, I think, all right—and even important. It's how it's reported, and thus interpreted, that is the problem. Unfortunately, 72 percent is 72 percent, whether it's a firm 72 percent or a 72 percent that can easily be pushed down to 32 percent; 72 percent is 72 percent, whether it's based on extensive knowledge or no more knowledge than is trans-

mitted in the wording of the question that produced the 72 percent figure.

I've also decided that having 50-word preambles to 20-word questions is legitimate and useful—not to *measure* public opinion, but to determine what opinion would be likely to be *if* people had the information contained in the preamble, or *when* they come to have that information.

Rather than deliver a definitive list of “dos and don'ts” as a means of improving our craft, I have tried to make three general points:

1. Let's emphasize the art part of opinion research more. That's the neglected side, that's the side that produces most of the error. Let's think about the art side more, let's talk about it more, let's experiment with it more.

2. Let's not fall into the trap of establishing standards of performance. Nothing could be better calculated to stifle creativity, to limit knowledge, to increase costs, and to minimize progress in the field—even if it raised the quality of the *worst* research to a somewhat higher level.

3. Let's push for disclosure—more disclosure than the present code calls for. Disclosure as to what alternate question wordings produce. Disclosure of knowledge levels as well as opinions. Disclosure of *whatever is pertinent* to a proper understanding of public opinion.

In closing, let me say that if the Opinion Research Fairy were to grant me just one wish this year, it would be to require everyone in the news media, and particularly the broadcast media, to eliminate their meaningless—in fact, highly misleading—error statement, and in its place substitute the wording of the question on which the reported results are based. There is no doubt in my mind that question wording has more to do with the results obtained than sampling error does. Nor is there any doubt in my mind that the lay public is better able to evaluate the wording of a question, and hence the result the question produces, than it is to understand the true meaning of an error statement.