who chose "moral values" as the issue that mattered most to them and those who did not: "What comes to mind when you think about moral values." It is true that traditional quantitative analysis is the major way in which we try to understand survey data, yet if our goal is to appreciate the meaning of answers to questions (and the meaning of questions to respondents), we will often wish to follow closed response choices by asking those we interview Why they answered as they did and then draw on what they say as an additional form of evidence when interpreting results. Bridging experience-near and experience-distant concepts can prove useful in ways that often cannot be anticipated.

— 4 —

Artifacts Are in the Mind of the Beholder

Who asks a question and where it occurs in a stream of questions can have a profound effect on the answers people give. We consider both kinds of context in this chapter, starting with the effects of one question on another question, and then turning to effects that involve interviewers and interviewing. Although context effects are sometimes thought of as artifacts, we are not dealing with error in the usual sense of that term, but with how the meaning of a question depends on the context in which it occurs. It is important to take this variation in meaning into account when planning or interpreting a poll or survey. In the end, context effects offer opportunities for deeper understanding of both responses and respondents.[1]

A Dawning Recognition of Question Context Effects

The possibility that questionnaire context (question order) might influence answers was recognized in the earliest days of polls and surveys but was seldom considered a major problem. In their broad review of response effects in surveys, Sudman and Bradburn (1974) concluded tentatively that the order of questions "has by itself . . . a negligible effect for attitudinal items" (p. 33). They referred also to a careful empirical study by Bradburn and Mason (1964), which stated as "its major conclusion . . . that responses to questions such as those tested—questions of self-report and self-evaluation—are relatively unaffected by order of presentation" (p. 57). At the same time, Bradburn and Mason noted mixed evidence from reports by others and acknowledged the impossibility of generalizing from their research "with any degree of confidence to other situations" (p. 61).

I was less cautious in a 1974 oral presentation, stating: "What strikes me most . . . is the extent to which respondents apparently consider each question in and of itself, without much attention to the earlier questions presented to them. The well-managed survey interview is more like a slide show than a motion picture, with each item viewed quite apart from what preceded or is to succeed it" (Schuman 1992, p. 5). The reason for my belief was that very few such effects had been reported in the survey literature at that point, and those few came partly from the Dark Ages of quota sampling, inattention to significance testing (which, whatever its limitations, introduces a certain amount of self-discipline into our endless search for positive results), and often inadequate reporting of basic data as well.[2]

These early conclusions by both myself and others remain true today in the sense that subsequent examinations of total survey questionnaires have not found frequent effects attributable to question order (Schuman and Presser ([1981] 1996, pp. 26–27; Smith 1991a).[3] At the same time, my own statement was clearly wrong in the sense that we now realize there are specific types of context effects that are substantial, reliable, and theoretically meaningful. One of the most interesting of these was discovered long ago, in the early days of the Cold War, but was missed by most survey and polling researchers because it was mentioned only in passing in a chapter that Hyman and Sheatsley (1950) wrote on American public opinion for an audience of educators. This effect has taken on a good deal more importance in recent years.[4]

The Norm of Reciprocity in Surveys as in Life

When each of the two questions shown in the top panel of Table 4.1 was asked first to half a national sample in 1948, Hyman and Sheatsley found that almost all Americans favored having American reporters obtain news from within the Soviet Union, but most were opposed to allowing foreign Communist reporters the same privilege within the United States.[5] When asked in sequence, however, the two questions—in either order—had a dramatic effect on answers to whichever one came second. Support for Communist reporters entering the United States rose sharply in the wake of the question about Americans reporting from Russia, and the results for the latter fell sharply when it was asked after the question about foreign Communists reporting from America. I was initially skeptical of the large effect

alluded to by Hyman and Sheatsley (1950), but when we replicated the sequence more than three decades later in 1980, the new results proved highly reliable, as shown in the bottom panel and footnote of Table 4.1 (Schuman and Presser 1981). From the standpoint of interpretation, the effect seemed consistent with what has come to be called the norm of reciprocity. Although not the only important type of context effect recognized today, it takes on broader theoretical significance because it connects the question–answer process to phenomena that are of interest to a range of sciences from anthropology to zoology.

Table 4.1 Percent yes to questions on American and Communist reporters

	Question order		Difference
	1st	2nd	
Americans or Russians (1948)			
"Do you think a Communist country like Russia should let American newspaper reporters come in and send back to America the news as they see it?"	90% (635)	66% (567)	−24%
"Do you think the United States should let Communist newspaper reporters from other countries come in here and send back to their papers the news as they see it?"	36% (581)	73% (635)	+37%
Americans or Russians (Replication, 1980)			
"Do you think a Communist country like Russia should let American newspaper reporters come in and send back to America the news as they see it?"	82% (331)	64% (336)	−18%
"Do you think the United States should let Communist newspaper reporters from other countries come in here and send back to their papers the news as they see it?"	55% (342)	75% (335)	+20%

Sources: National Opinion Research Center Survey (1948); University of Michigan's Survey of Consumer Attitudes (May and July 1980 combined).

Notes: Base Ns are shown in parentheses. Small DK percentages are omitted. Using chi square tests, with df = 1, each of the four differences is highly significant ($p < .001$).

The Nature and Reach of the Norm of Reciprocity

According to Gouldner (1960), the norm of reciprocity refers to the "mutually contingent exchange of benefits between two or more units" (p. 164). More simply, it is the expectation that if A does something that benefits B, B is expected to do something in return of approximately equivalent value. A and B can be individuals or groups, and if groups they can be as small as couples or as large as entire nations. Because there is a sense of obligation to repay a benefit, one felt by both parties involved, as well as perceptible to outside observers, the concept of "norm" in the sense of "prescription" applies. An implicit feature of the exchange is some temporal delay between the giving of a benefit and its repayment, for otherwise the exchange would be an ordinary economic transaction. Thus, Hauser (2006) treats the ability to defer gratification as essential when an exchange is governed by the norm of reciprocity. Furthermore, there must be at least the possibility of no repayment on the part of the recipient, for otherwise the normative—obligatory—aspect would be lacking.

Each discipline approaches the norm of reciprocity from its own perspective and with its own issues. Within anthropology, Malinowski (1926) wrote about the chain of obligations for both practical and ceremonial exchanges between Trobriand fisherman and agriculturalists, while the sociologist Marcel Mauss (1954) drew on ethnographic reports to describe extravagant ritual feasts (the potlatch) where competition for prestige was the central element in both giving and repaying. Sociologists like Simmel (1950) and Gouldner (1960) have stressed the universality of the norm and its importance for social cohesion, while Homans (1950) and Blau (1967) paid special attention to interpersonal relations and stressed the role of social approval more than material gifts. Developmental psychologists like Piaget (1932) have attempted to identify the age at which the norm of reciprocity first influences the behavior of children, especially in their relations to peers (e.g., Rottenberg and Mann 1986).

In addition to the sociological and related approaches, evolutionary biologists like Trivers (1971) have tried to explain how reciprocity can occur at all, in the face of what is frequently seen as the Darwinian emphasis on "fierce competition between individuals [which should] reward only selfish behavior" (Nowak 2006, p. 1560). However, de Waal (2006), quotes from *The Descent of Man* to argue that Darwin believed social instincts shown by humans to be found in some form in other animals, and de Waal himself reports having observed reciprocal actions between chimpanzees that included a clear time delay. Axelrod (1984) developed a widely cited theory of how reciprocal responses among people can lead to long-term cooperation in iterated prisoner dilemma games, and others (e.g., Fehr and Henrich 2003) have tried to explain how reciprocity can appear even in single encounters. The norm also operates in small routinized ways, as when two social scientists sent Christmas cards to a sample of people listed in Polk Directories and found that 20 percent reciprocated by replying with their own card to the unknown sender (Kunz and Woolcott 1976). Indeed, surveys often attempt to make use of the norm of reciprocity by adding either money or a nonmonetary gift to an initial request for an interview, and these actions have been shown repeatedly to improve willingness to participate (Singer 2002).

A Simple View of Context Effects Attributable to the Norm of Reciprocity

Here our concern is with how the norm of reciprocity affects answers to survey questions. It does this not by influencing each of the parties described in the question—for example, Russians and Americans—but instead by influencing the responses given by one of the parties (Americans, in these experiments) when the relevance of the norm is highlighted by the order of questions. This is an important difference from reciprocal actions in real life—a difference that we will need to consider at a later point.[6]

There is no other past experiment quite as clear in its invocation of the norm of reciprocity as the reporters questions, but three further context experiments, also from the 1940s, draw on awareness of an obligation for reciprocity in responding to an attitude question. The norm itself shades into more general issues of equitable treatment or "fairness," as recognized by Hauser (2006) and by Walster, Walster, and Berscheid (1978). Thus, each of the pairs of questions in Table 4.2 should alert respondents to take account not only of their own preference on the issue, but also of the claims for equal treatment on the part of someone holding an opposing position. The three experiments concern both foreign and domestic issues:

1. *Friend or Foe*
The earliest instance of a context effect attributable to the norm of reciprocity was reported by Rugg and Cantril (1944) for a split-sample experiment carried out by Gallup in 1939 at the outbreak of World

Table 4.2 Percent yes to each question in three context experiments

	Question order		Difference
	1st	2nd	(2nd – 1st)
1. *Friend or Foe* (asked in 1939; from Rugg and Cantril, 1944)[a]			
"Should the United States permit its citizens to join the French and British armies?"	49% (1415)	43% (1570)	–6%
"Should the United States permit its citizens to join the German army?"	23% (1495)	34% (1431)	+11%
2. *Strikes or Lockouts* (Link, 1946)[b]			
"Do you believe that workers and unions have the right to strike when wages and working conditions don't suit them?"	66%	62%	–4%
"Do you believe that businessmen have a right to shut down their factories and stores when labor conditions and profits don't suit them?"	47%	52%	+5%
3. *Political Contributions by Corporations or Labor* (Gallup 1947)[c]			
"Do you think labor unions should be permitted to spend labor funds (money) to help elect or defeat candidates for political offices?"	23% (1376)	16% (1313)	–7%
"Do you think business corporations should be permitted to spend corporation funds (money) to help elect or defeat candidates for political offices?"	14% (1320)	24% (1362)	+10%

Sources: Data for the experiments 1 and 3 are obtainable from the Roper Center. Results for experiment 2 are from Link (1946), with the total sample size said to be 5,000, divided in half for the experiment but with exact N's not given.

Notes: Base N's are shown in parentheses.

a. Each difference is significant at $p < .005$. Percentages differ from those in Cantril (1944) because DK responses are removed here.

b. Exact Ns on which percentages are based were not given, but total N was reported as 5,000, with approximately 2,500 on each of two forms. Each difference is said to be significant at $p < .01$ (Link 1946). Author was not able to locate the original data, which may not have been archived.

c. Each difference is significant at $p < .001$.

War II. One question to a national sample asked whether American citizens should be allowed to join the British or French armies; the other question asked about Americans joining the German army. When each question appeared in the first location, respondents were more favorable to Americans joining the Allied forces than the German army by 2 to 1, but when the same questions were asked in the second location the ratio shrunk to just 1.3 to 1.[7]

2. *Strikes or Lockouts*

Link (1946) asked one question about the rights of workers to strike and another question about the rights of businessmen to lock out workers. Support for strikes was 19 percentage points greater than support for lockouts when each was asked in the first position, but the difference decreased to 9 percentage points when each was moved to the second position.[8] The differences are reported to be significant at $p < .01$.

3. *Political Contributions by Corporations or Labor Unions*

Gallup carried out this experiment in 1947, but apparently did not publish it at that time. As with the Link experiment, there was more support (by 10 percentage points) for the union side than for the corporation side when each question was asked first, but support for the corporate side became greater and support for unions lessened when the questions were in the second position. The differences are significant at $p < .001$.[9]

The three experiments, along with the original Hyman and Sheatsley experiment on American and Russian reporters, were all carried out within several years of one another in the middle and late1940s, but none of the investigators showed evidence of being aware of the other experiments, nor were the several results ever brought together as part of a more general treatment of context effects, or of reciprocity or fairness. Each experiment appears to have been done primarily, or perhaps only, to prevent misinterpretation of substantive results because of question order, without explicit recognition that the context effect comes from the intrinsic connection between related actions and thus omitting one question does not eliminate but simply changes the effect. In none of the four cases did the nature of the context effects become a focus for further experimentation or analysis.[10]

Recent treatments of context effects accept the empirical findings reported in Tables 4.1 and 4.2 as trustworthy, and the explanation in terms of the norm of reciprocity as persuasive (Sudman, Bradburn, and Schwarz

1996; Tourangeau and Rasinski 1988). The basic result, especially as shown by the reporters experiment, has become a standard example of the importance of context in surveys and of a case where both the data and its interpretation are unproblematic (e.g., Weisberg 2005). However, if we go beyond the comparisons of the distributions shown in Tables 4.1 and 4.2, and add further evidence and analysis, complexities arise that point to changes in the way we should view both the data and their meaning.

A More Complex View of Context Effects and the Norm of Reciprocity

INTRODUCING TIME. The 1980 data on the reporter's experiment provided not only a replication, but a replication after a considerable period of time, thus allowing real change along with a test of the reliability of the context effect. Looking more closely at the results for the questions in Table 4.1, we can note that the differences attributable to the norm of reciprocity were not as great in 1980 as in 1948, though if statistically significant effects due to context are one's sole concern, the possible change over time might easily be ignored. This would be even truer if the results were based on two convenience samples, for example, students from a class or volunteers. With probability sampling from the same national population, however, the change encourages fuller consideration of "time" as an important variable. Moreover, the decrease in the size of each context effect between 1948 and 1980 is highly reliable: for the United States allowing Communist reporters item: $\chi^2 = 10.1$, $df = 1$, $p = .001$; for Russia allowing American reporters item: $\chi^2 = 5.8$, $df = 1$, $p = .02$.

Suppose now that we consider the possibility of attitude change toward Communist and American reporters continuing to move in the same direction, as indicated in Table 4.1. This is not simply idle speculation: it might really have happened if Gorbachev's attempt to reform the Communist system in Russia had been more successful, not only internally but in winding down the Cold War as well. The effect of such further change is easiest to observe if we focus on the question about allowing Communist reporters into the United States when it is asked first. Suppose by, say, the year 2000, with the Cold War much diminished, the response "Yes" had risen another 20 percent to 75 percent, as shown in the hypothetical Table 4.3, and in addition the result when the question is asked second did not change—as it had not changed between 1948 and 1980—then the context effect would

have disappeared entirely. In other words, what Tourangeau, Rips, and Rasinski (2000, p. 211) regard as the largest context effect ever discovered in a national survey would no longer exist! It is less easy to imagine that support for American reporters to be let into Russia would have declined enough to erase that context effect completely, but its size and statistical significance might well have faded considerably. More generally, the context effect in the case of the reporter's questions cannot be divorced from attitudes toward the object itself.[11]

Strack and Martin (1987) have proposed that because cognitive scientists are interested in knowledge about processes, rather than knowledge about content, it is unnecessary for them to do experiments in sample surveys of general populations, with all the additional costs and time that such surveys require. Moreover, they go further and argue that it is better to use homogeneous convenience samples so that there is not unwanted variation in responses due to relations between content and social background variables such as education. But in the case of the norm of reciprocity and the reporter's items, process and content are tied together, and I'm not persuaded that it is possible to study the one and ignore the other completely.

Most psychologists who do laboratory-type experiments seldom think of including long-term social change as an important variable. However, if they did, how could they tell that a difference over time that they discover is not due to the particular convenience sample of students used at each point? And even if they controlled sampling well enough to feel confident of real change, how could they know that what was going on with their college students was also happening with the larger public? Studying change calls for probability samples from a well-defined population, and generalizes only to that population (Sears 1986).[12] This is why sample surveys are done when there is interest in studying the content of attitudes. They may be

Table 4.3 Percent yes about Communist reporters: Hypothetical year 2000

	Question order	
	1st	2nd
"Do you think the United States should let Communist newspaper reporters from other countries come in here and send back to their papers the news as they see it?"	75% (581)	75% (635)

Notes: Percentages are hypothetical, as described in the text. Base N's in parentheses are from 1948 data for the Russia question.

crucial as well when our concern is with processes such as context effects, and perhaps other response effects as well, especially as our time horizon begins to extend far enough to alter the way such effects are expressed in the intrinsically social medium of language.

A CONSTRUCT REPLICATION. The findings of context effects attributable to the norm of reciprocity in Tables 4.1 and 4.2 were all reported after the fact, and with some possibility that they were unplanned even by the investigators, other than as a way of preventing possible influence on their substantive concerns.[13] If the operation of the norm of reciprocity is as clear as now believed, we should be able to design a new experiment that will yield an entirely novel context effect.

National trade barriers seemed to provide a particularly suitable form of construct replication, for trade negotiations are often stated explicitly within a framework of reciprocity.[14] Therefore, Schuman and Ludwig (1983) constructed the order experiment in 1982 that is shown in Table 4.4, and predicted an approximately symmetrical context effect similar to the effects shown for earlier experiments in Tables 4.1 and 4.2. Indeed, the prediction seemed so plausible that reviews of context effects have added its supposedly successful outcome to past evidence in support of context effects due to the norm of reciprocity (e.g., Sudman, Bradburn, and Schwarz 1996; Tourangeau, Rips, and Rasinski 2000). But let's look closely at the actual results.

Not surprisingly, when each question was asked first, many more Americans supported trade restrictions by the United States (76 percent) than supported trade restrictions by Japan (48 percent); see Table 4.4. Thus, self-interest prevails when the norm of reciprocity is not made explicit, though it is interesting to note that nearly half the sample agreed to Japan's use of trade restrictions even when that question came first. This is an indication that national self-interest can sometimes be put aside, a finding we return to at a later point.

In accord with our main hypothesis, there is a large and significant context effect on Japan's use of trade limits when that question comes after the question about trade limits by the United States: allowing trade barriers by Japan increases from 48 percent to 70 percent, reaching almost the level of support given to the preceding item about allowing the United States to set barriers against products from Japan (76 percent). Apparently the norm of reciprocity leads nearly a quarter of the American sample to allow Japan to do something that these Americans would otherwise oppose.

Table 4.4 Context effects on American and Japanese trade restrictions

	Question order			
	1st	2nd	Difference	
"Do you think that the American government should be allowed to set limits on how much Japanese industry can sell in the United States?" Yes, allowed:	76% (195)	71% (186)	−5%	$\chi^2=1.2$, 1 df, p=n.s.
"Do you think that the Japanese government should be allowed to set limits on how much American industry Can sell in Japan?" Yes, allowed:	48% (186)	70% (195)	+22%	$\chi^2=20.0$, 1 df, p=.001

Source: University of Michigan's Survey of Consumer Attitudes, February, 1982.
Note: Base N's are shown in parentheses.

There is also a slight decline in support for U.S. trade restrictions when that question appears in second position, but contrary to our hypothesis the effect is quite small (5 percent) relative to the effect in the other direction, and unlike differences in earlier tables, it does not approach statistical significance ($p>.25$). Furthermore, using a telephone survey of American college students, Bishop et al. (1988) report similar findings: a clear context effect on allowing Japan to have trade limits when it follows the American item ($p<.01$), but no reliable effect on reducing U.S. trade barriers when that question follows the Japan question ($p>.25$). Thus, we find that in this case the pull of self-interest is so great that few, if any, Americans would change their position on the U.S. right to restrict imports from Japan, even after indicating unwillingness to allow Japan to restrict American imports. It is likely that economic problems arising out of Japanese–U.S. trade relations at that point in time were perceived by a large number of Americans almost entirely in terms of competition for the U.S. market.[15]

In sum, this new experiment created an asymmetrical context effect that forces us to modify the previous implicit assumption that the norm of reciprocity always has similar implications in both directions in these experiments. We can also now review the pairs of questions in Tables 4.1 and 4.2,

and note that for all of the comparisons, the more popular side, as indicated by the larger percentage of Yes responses in first position, has the smaller reduction as a result of the norm of reciprocity, though sometimes the difference is tiny. The first positions also make intuitive sense because a plurality of Americans seems likely to have favored these views in the 1940s and 1980s. Thus, the trade restriction results may simply be an extreme example of a more general competition between the principle of reciprocity and personal preference.

These conclusions provide compelling evidence that the norm of reciprocity in surveys should not be regarded as an abstract principle that can transcend personal preference and self-interest in concrete situations. The results also point up an important difference between survey-based experiments on reciprocity and what goes on in real life where each party in a reciprocal interaction is almost certain to try hard to prevent being disadvantaged in the exchange. Our experiments should not be treated as reproducing exactly what happens in nonsurvey situations—just as dropping two objects having different weights in a vacuum is not the same as dropping them through currents of ordinary air. Yet survey-based experiments do have the advantage of allowing us to explore further issues about reciprocity effects.

Carriers of the Norm of Reciprocity

Narrow self-interest can hardly account for the fact that nearly half the 1982 sample favored allowing Japan to create trade restrictions even when that question was asked first, and in addition a quarter would oppose American trade barriers when it came first.[16] Thus, even when the norm of reciprocity is not made salient, some respondents answer with the norm in mind. To test this assumption, we made use of a separate small national sample of 41 respondents in March 1982, asking only the question about Japanese trade limits, and then for those who said that Japan should be allowed to set limits, we asked: "Could you tell me why you feel that way?" Of the 24 respondents who said Yes, nearly three-quarters (17) gave a reason clearly involving reciprocity. For example, one person said: "We set limits, so they should be allowed to set limits . . . it's only fair." Thus, for some respondents the norm of reciprocity comes to mind without prompting when a question to which it is relevant is posed.[17]

A plausible hypothesis is that it is more knowledgeable and sophisticated respondents who carry with them awareness of the relevance of the

reciprocity norm, whereas others recognize the norm only when it is made salient by context, and still others are not affected by the norm at all. Educational attainment can provide an indicator of both knowledge and cognitive sophistication, and might therefore be expected to interact with context meaningfully. Table 4.5 presents results by education for the trade restrictions experiment, with income controlled, and the findings are consistent with this expectation. When a trade question is asked first, more educated respondents are less likely to support U.S. restrictions on Japanese goods, but more likely to support Japan's right to restrict American goods, than are those with less education. Furthermore, the context effect of the norm of reciprocity is greatest for those with the least education.[18] Thus, it is not that the force of the norm of reciprocity necessarily differs for those at different educational levels on this question, but rather that it is more salient ("chronically present") to those with more education, and requires context to bring it to mind for those with less education.

If the overriding factor in determining responses is the power of the norm of reciprocity, we would expect similar results for the three additional experiments where data are available that include measures of both educa-

Table 4.5 Percent yes on trade restrictions, by education and question context

	Education		
Trade Restrictions Context Effect	0–11	12	13+
Yes, Trade Restrictions by Japan			
Order: Restrictions by Japan 1st	39% (31)	45% (69)	56% (86)
Order: Restrictions by Japan 2nd	86% (29)	67% (69)	67% (100)
Odds ratio:	.10	.41	.62
Response × order × education: linear $\chi^2 = 6.4$, df = 1, $p < .01$			
Yes, Trade Restrictions by U.S.			
Order: Limits by U.S. 1st	97% (30)	75% (69)	68% (101)
Order: Limits by U.S. 2nd	67% (33)	74% (69)	70% (87)
Odds ratio:	.07	.93	1.09
Response × order × education: linear $\chi^2 = 6.0$, df = 1, $p < .02$			

Source: University of Michigan's Survey of Consumer Attitudes, February, 1982.
Note: Base Ns are shown in parentheses.

tion and income: the two reporters experiments from 1948 and 1980 and the experiment on union and corporation political contributions. However, in Table 4.6 none of these shows a three-way interaction similar to that for the trade restriction experiment, as indicated by the odds ratios within each of the tables and also by formal tests of interaction between each question, context, and education, with income controlled.[19]

Yet the reporter experiments do show important regularities that are much the same in both 1948 and 1980. In all eight tests in Table 4.6, as the regression coefficients indicate, there is a highly reliable relation between more education and more support for allowing foreign Communist reporters into the United States and for allowing American reporters into Russia. Based on other evidence (e.g., Stouffer 1955), it seems likely that this is a result of greater support for freedom of expression and greater tolerance for nonconformity by more educated Americans. Furthermore, since this result occurs even when a question is in second position and thus subject to influence by the norm of reciprocity, it is apparently too strong to permit the kind of reversal that appeared in Table 4.5 for trade restrictions.

At the same time, we can see that the norm of reciprocity does continue to be effective within educational categories: each question in second position shows the same direction of influence as appeared in Tables 4.1, 4.2, and 4.4 for the total samples. For example, among those with 0 to 11 years of education in 1948, agreeing to allow Communist reporters into the United States rises from 25 percent when the question comes first to 63 percent when it follows the question about American reporters going to Russia. By the same token, among those with 0 to 11years of education, insistence on American reporters being allowed into Russia drops from 84 percent to 57 percent when that question follows the question on admitting Communist reporters into the United States. There are twelve possible comparisons in Table 4.6, and all twelve are in the direction predicted by the norm of reciprocity.

When we turn to the final experiment on political contributions, we again find that within each educational category the results are in the direction predicted by the reciprocity norm. For example, considering only first positions, there is more support within each of the three educational categories for contributions by unions than by corporations. Given that starting point as an indicator that unions were more favored at that point in time than corporations with regard to making political contributions, we expect support for union contributions to go down when the question is in second position and for support for corporation contributions to rise when it is in

Table 4.6 Percent yes on reporters, by education and question context

	Education		
A. Reporters (1948)	0–11	12	13+
U.S. Allow Communist Reporters			
Order: U.S. 1st	25% (337)	40% (129)	66% (113)
Order: U.S. 2nd	63% (310)	79% (169)	87% (153)
Odds ratio:	.50	.44	.54
U.S. 1st: b=−.73 SE=.13 p<.001; U.S. 2nd: b=−.57 SE=.14 p=.001.			
Russia Allows U.S. Reporters			
Order: Russia 1st	84% (311)	93% (168)	97% (153)
Order: Russia 2nd	57% (324)	69% (125)	87% (116)
Odds ratio:	.79	.45	.62
Russia 1st: b=−.74 SE=.23 p=.001; Russia 2nd: b=−.58 SE=.15 p=.000.			
B. Reporters (1980)			
U.S. Allow Communist Reporters			
Order: U.S. 1st	24% (70)	43% (122)	78% (141)
Order: U.S. 2nd	54% (69)	76% (105)	84% (153)
Odds ratio:	.50	.44	.54
U.S. 1st b=−1.17 SE=.18, p=.000; U.S. 2nd: b=−0.58 SE=.18 p=.001.			
Russia Allows U.S. Reporters			
Order: Russia 1st	68% (66)	84% (104)	88% (153)
Order: Russia 2nd	35% (66)	58% (123)	83% (138)
Odds ratio:	.79	.45	.62
Russia 1st: b=−.53 SE=.20 p=.007; Russia 2nd: b=−1.00 SE=.18 p=.000.			

Sources: Data for 1948 reporters experiment are obtainable from the Roper Center; data for 1980 reporters experiment are from the University of Michigan's Survey of Consumer Attitudes.

Note: Base Ns are shown in parentheses.

second position. Each of these predictions is confirmed in Table 4.7 where there are six comparisons to consider.

However, there is not consistent evidence for education as a predictor of responses either in the interactive sense that was true for the trade restriction experiment, or in the simple correlational sense that held for the 1948

Table 4.7 Percent yes on political contributions

	Education		
Political Contributions	0–11	12	13+
Yes, Contributions by Corporations			
Order: Corporations 1st	13% (655)	12% (310)	17% (312)
Order: Corporations 2nd	23% (661)	24% (362)	27% (295)
Odds ratio:	.50	.44	.54
Corp. 1st: b=–.17 SE=.11 p=n.s.	Corp. 2nd: b=–.16 SE=.10 p=.09		
Yes, Contributions by Unions			
Order: Unions 1st	21% (671)	23% (367)	27% (294)
Order: Unions 2nd	17% (648)	12% (311)	19% (312)
Odds ratio:	.79	.45	.62
Union 1st: b=–.28 SE=.10 p=.003	Union 2nd: b=–.11 SE=.11 p=n.s.		

Source: Gallup 1947 (Data are obtainable from the Roper Center for Public Opinion Research).

and 1980 reporters experiments. More generally, each of the three subjects about which we have data for analysis (trade restrictions, reporting by a foreigner, and political contributions) shows a distinctive overall relation between respondent education and the norm of reciprocity, though in each case the efficacy of the norm within educational categories is clear and consistent. A full synthesis of these results for education is not at hand, and thus the familiar refrain of a need for more research—and thinking—is appropriate at this point on this problem.

Conclusions about the Norm of Reciprocity and Context Effect

Implications for Reciprocity in Life

The norm of reciprocity is typically discussed as if symmetrical: A does something that benefits B, and B is expected to do something in return of approximately equivalent value, and so on. However, when looked at from the perspective of an individual evaluating both directions of a reciprocal interaction, many respondents see it as more appealing in one direction than in the other, and this may often be the case with reciprocity between two individuals or two groups, with disagreement or conflict resulting.

A related possibility worth testing in the future is that a benefit may be regarded as obligatory to return, but not necessarily in equal amount or quality. This was nicely captured by a cartoon printed in the *Ladies' Home Journal.* As they are leaving a house where they have obviously just been dinner guests, a wife remarks to her husband: "Eggplant casserole doesn't pay back for prime rib."[20] A dinner has been reciprocated, but not in a way judged to be equivalent. There may certainly be such instances in ordinary life where the reciprocation seems inadequate, and it would be useful if a future experiment attempted to measure actions along an ordinal or interval scale. Perhaps what we tap only in terms of percentages of individuals who do or do not favor reciprocity in particular situations would appear differently if degree of support or opposition had been assessed. In much of the social science literature on reciprocity the same assumption of a dichotomy is made.[21]

There are also ways in which close relationships can modify the expression of reciprocity in important respects. In Whyte's classic study of *Street Corner Society* (1943), the author writes:

> Once Doc asked me to do something for him, and I said that he had done so much for me that I welcomed the chance to reciprocate. He objected: "I don't want it that way. I want you to do this for me because you're my friend. That's all." (p. 256)

Not all relationships call for or can tolerate reciprocity that is too blatant. Moreover, any relationship can experience changes in attitudes that make reciprocity difficult to expect or to assess, unless the attitude change itself is taken into account. This applies also to relationships between groups and organizations, so we should not assume that reciprocity once observed is sure to continue cycling on indefinitely.

We were able to explore the influence of the reciprocity norm in relation to educational level in three quite different spheres: trade barriers between nations, support for reporters being admitted into countries seen as enemies, and political contributions by organizations (unions and corporations) with competing agendas. Just as the efficacy of the norm varied over time, its nature seemed to vary across spheres in relation to education and perhaps to other social factors not tested. Much (though not all) of the research on reciprocity approaches the norm as an abstract force, unmodified by other social variables and social settings, but the results of the present analysis suggests that this is too rarified a view.

Implications for Survey Research

There are important implications for surveys from our study of context effects due to the reciprocity norm. First, we must recognize not only that such effects exist and are important, but that they are due to the integral nature of certain attitudes toward other actors. Reciprocity can have a role in survey questions whether it is implicit for some respondents, is made salient to others by question order, or is either missed or rejected entirely by still others. Similar conclusions come from context experiments on how respondents answer questions about general life satisfaction and satisfaction in specific spheres such as marriage: there are effects of one kind or another no matter what particular questions are included or how they are ordered (Schwarz, Strack, and Mai 1991). We are not dealing with a form of "survey error" in any simple sense.

Second, Tourangeau, Rips, and Rasinski (2000) have classified context effects in terms of a four-stage psychological process that most respondents are assumed to go through when answering an attitude question: first, they must interpret the question; second, they try to retrieve relevant beliefs and feelings; third, using what they are able to retrieve they make a judgment about a desired response; and finally they edit the judgment to fit the alternatives offered and their own need for self-presentation. Sudman, Bradburn, and Schwarz (1996) use a similar classification. However, it is not clear how this works with particular context effects. Tourangeau and Rasinski (1999, p. 306) place the effect for the reporters items at the "judgment" stage, but also note that it may possibly affect retrieval ("what respondents consider in making their judgments"), and presumably the same would be true for the trade limits effect. In addition, it seems possible that some respondents arrive at their final response at the reporting stage, once they see the need to appear consistent in the eyes of their interviewer and indeed to themselves. With a little ingenuity, one might even manage to locate reciprocity context effects at the comprehension stage. In addition, to complicate matters more, it is entirely possible that for some respondents the effect is mainly at one stage, for others mainly at a different stage. Thus, rather than the four components of the response process providing a simple a priori solution "by the book," the attempt to apply the classification to real context effects calls for evidence of what actually goes on with different types of effects and perhaps different kinds of individuals.

Third, it is always useful to attempt three types of replication in a study of context effects, as indeed of response effects generally. There is the initial need for literal replication to make certain that what looks to be a reliable finding is really reliable, because significance tests and similar measures from a single survey are seldom if ever sufficient for that purpose. Next, replication over time is valuable to determine whether a conclusion needs to be changed because attitudes have shifted in some way that alters earlier findings in important, even decisive, respects. "Chance" is not the only source of a failure to replicate an earlier result, as we will see even more clearly in Chapter 6. Beyond both types of literal replication, construct replication is essential to distinguish findings due to specific content from findings due to constructs that the content is assumed to represent. We saw that what seemed like a reasonable extrapolation from one experiment (reporters) to another (trade barriers) did not in fact produce the expected results, and the same turned out to be true when we looked at the impact of education on different applications of the norm. There has been a tendency by those writing about response effects to assume on the basis of a single experiment (e.g., the 1948 Hyman and Sheatsley reporter questions) that effects due to the norm of reciprocity are simple and straightforward, but the analysis in this chapter makes it clear that that is not the case.

Fourth, although much research on context effects can be done with convenience samples such as students, at some points it is important to work with probability samples of a well-defined and heterogeneous population. On the one hand, this is important if replication over time is to be carried out to assess change in the nature of context effects. Such replication calls for the population to be the same at different time points, or at least adjustable to take account of differences. On the other hand, the inclusion of important variables like education as part of the analysis is often accomplished most adequately by working with natural populations, for example, the population of American adults. This is of course expensive and time-consuming, but needed nonetheless.

Finally, because important context effects in surveys are always possible, but are also thought to be uncommon, we have another reason for all major surveys to include split-sample question order experiments, especially where attitudes central to an investigation are being studied. This is much like the efforts in the 1940s that may have been done mainly for substantive reasons, but resulted in the first clear findings of important context effects

in surveys. Both substantive and methodological goals are addressed by continuing to construct experiments that identify new context effects.[22] Furthermore, it is also desirable to include open-ended Why follow-up inquiries, as discussed in Chapter 3. The open-ended inquiry about trade restrictions in the present study showed that some respondents had the norm of reciprocity in mind even when it was not made salient by question order. The use of experimentation together with open questioning, along with replications across time and analysis of associations with education and other important social variables, brings to bear the strengths of four different methods in pursuit of fuller understanding of the meaning to be found in questionnaire context.

Postscript on Other Context and Wording Effects

Effects due to reciprocity ordinarily involve two questions of equal specificity, with influence assumed to flow in both directions and usually hypothesized to create greater consistency between responses to the questions. Most other context effects discovered thus far involve what Schuman and Presser ([1981] 1996) called "Part-Whole relations," where one question is more general (the Whole) and is intended to contain, summarize, or imply questions that are more specific (e.g., overall happiness and happiness with one's marriage, or support for legalized abortion generally and support for legalized abortion in the case of a fetus with serious defects). In these cases, context effects have been found for the specific question on the general one, but not the reverse. Furthermore, Part-Whole effects can produce either greater consistency or greater contrast in responses. Although some experimentation on these further types of effects was reported by Schuman and Presser, most of the important research has been developed by others. Starting points for what is now a large literature are Tourangeau and Rasinski (1988; see also Tourangeau, Rips, and Rasinski 2000) and Schwarz and Bless (1992; 2007), with the latter's "inclusion/exclusion" model providing probably the most productive theorizing at present.

Context effects often involve radical shifts in interpretations of words, as in the case of the Communist and American reporters: in one context attitudes toward objects (Communist and American reporters) are the focus, but the other context highlights the norm of reciprocity. This shift can produce different patterns of change over time, as shown in Figure 4.1a, which repeats graphically results from Table 4.1 for the United States letting in a

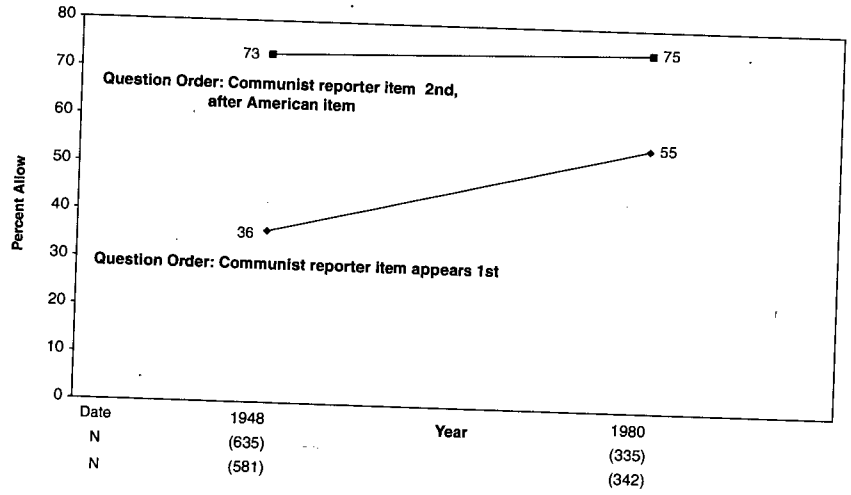

Figure 4.1a Question order

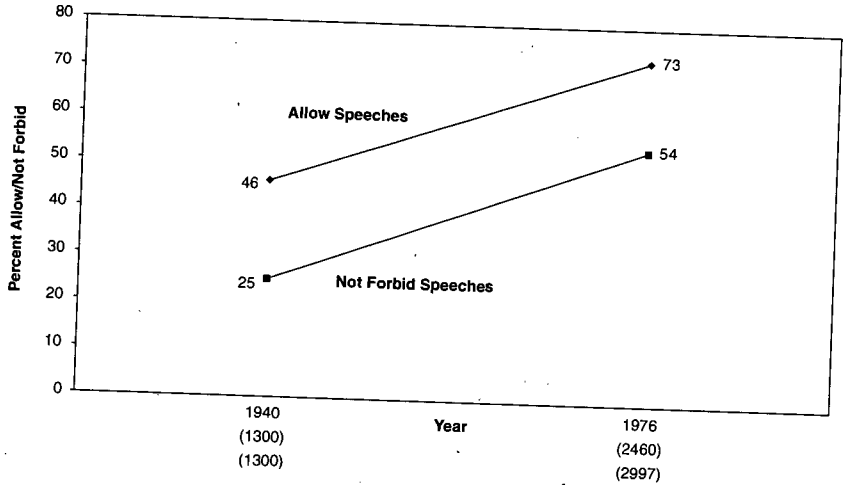

Figure 4.1b Allow versus Not Forbid unpopular speeches by year.
Source: Adapted from Schuman 2002.

Communist reporter.[23] However, in the case of some other types of response effects where a word within a question is altered, there may be a sharp difference at each point in time, but no difference in the nature (slope) of the association with time. This is true of the classic wording effect due to the difference between "allowing" an unpopular speech and "not forbidding" the

speech, as shown in Figure 4.1b for two time points (Rugg 1941; Schuman and Presser [1981] 1996). Not only visually, but also statistically, there is little sign of an interaction with year ($\chi^2 = 1.1$, df = 1, n.s.), even though the difference in wording has a clear impact at each time point.[24] The alteration of wording in this case evidently involves movement in degree of intensity along a single dimension, but not a radical shift in meaning. The difference between response effects that change meaning and those that do not is fundamental.

The Interviewer as Context

Interviewers are ordinarily expected to serve as transparent intermediaries, for the most part simply asking questions and recording answers. But knowledgeable survey researchers recognized almost from the start that responses to a survey can be affected by who asks the questions. During World War II, Stouffer and his associates in the Research Branch of the War Department compared answers of black soldiers to black interviewers with answers of a matched sample of black soldiers to white interviewers. The black interviewers "tended more than white interviewers to elicit responses reflecting racial protest, unfavorable attitudes toward the war, pessimistic views of postwar conditions, unfavorable reports on the Army, and manifestations of low personal esprit" (Stouffer, 1950, p. 721). Although the differences might possibly have been due to conscious or unconscious biasing behavior by the interviewers, Stouffer added a third sample of black respondents who filled out the same questionnaires in a classroom situation with a black administrator and found that it yielded answers similar to those given to black interviewers, thus providing evidence that it was the soldiers' perceptions and beliefs that influenced their responses, not the interviewers' bias in how they asked the questions or reacted to the answers. Of course, a fourth sample using a white administrator would have clarified the results still further, and more generally the issue of perceived sponsorship of a poll may be as important as who asks the questions (e.g., Presser, Blair, and Triplett 1992).

Similar results by race of interviewer were obtained by Hyman (1954) when he compared African Americans interviewed by black and by white interviewers in Memphis in 1942, but with the additional finding that significantly fewer race-of-interviewer differences occurred when the same design was repeated in New York City. This expanded the interpretation of

interviewer effects by emphasizing the importance of local norms in increasing or decreasing the strength of the effects. Hyman's research also added to evidence that interviewer effects in professionally managed surveys come more from respondent attempts to avoid saying things they think are likely to offend interviewers than from direct bias in words or other behavior on the part of the interviewers themselves.[25] Moreover, to the extent that interviewers' own behavior proved important, Hyman provided other evidence that this was not due to ideological bias but because the expectations interviewers hold sometimes leads them to misinterpret answers given by respondents. Thus, on both sides of the interviewer-respondent divide, effects appeared to be due to what the one side assumed—perhaps incorrectly—about the other.

In 1968 Jean Converse and I reviewed a large number of race-of-interviewer effects and noneffects on the answers black respondents gave in a face-to-face survey in metropolitan Detroit. We came to conclusions much like those of Stouffer and Hyman: "To accept a guest into your house and then proceed to explain that you neither trust nor feel friendly toward people of their race probably takes more *chutzpah* than the average respondent possesses" (Schuman and Converse 1971, p. 58). At the same time, we also speculated that with the rise of black militancy at that point in time, it was possible that some black respondents felt a need to exaggerate their anti-white attitudes when the interviewer was black, though no direct evidence of such a phenomenon could be presented based on the available data.[26]

Furthermore, some attitudes can be conceptualized as entirely situational, different but equally valid in different situations. For example, a seemingly neutral question in the 1968 survey asked respondents simply to name their "favorite actors and entertainers" and elicited more black names to black interviewers and more white names to white interviewers. In such cases it is possible that respondent perceptions of the interviewer "primes" racially congruent names without there being a deep-seated motive to please or avoid displeasing the interviewer. The effect is a large one and thus it would be valuable to determine what the correct explanation is.

We also drew the broader conclusion that interviewer effects should be regarded not simply as an artifact of surveys, but as a fact of life in America, with the same effects likely to occur in other spheres where blacks and whites interact. This conclusion in turn suggests that a change over time in race-of-interviewer effects could provide a sensitive barometer of important

shifts in the nature of black/white interaction in the United States. With possible future change in mind, we were candid in stressing the time point of our own research by including it in the title of the article: "The Effects of Black and White Interviewers on Black Responses in 1968." A similar measurement over time would be valuable for race-of-interviewer effects on white respondents, since a small survey-based experiment in 1971 indicated that whites were also more likely to express positive racial attitudes to black interviewers than to white interviewers on such issues as racial intermarriage (Hatchett and Schuman 1975–1976), probably for much the same reason of politeness. Unfortunately, later replications were never carried out of either effect to assess long-term change.[27]

Two limitations in most race-of-interviewer findings have been pointed out by Groves (1989). First, since observations are clustered within individual interviewers, calculations of standard errors should take this into account, rather than treating the individual cases as a simple random sample, as was done in my and other earlier studies. This change in calculations means that the differences are likely to be less significant than originally assumed. For the present writing I reanalyzed five items from the 1968 study, with individual interviewers treated as a random factor nested within race of interviewer (Dijkstra 1983), and the original conclusions are essentially unchanged. Furthermore, the effects found in 1968 for black respondents were generally replicated three years later in 1971 with a new sample of interviewers, as well as a new sample of respondents (Schuman and Hatchett 1974), increasing our confidence in the reliability of the original results.

However, a second difficulty with studies of race of interviewer is conceptual: when a physical attribute of interviewers is varied, it is difficult to determine clearly the source of the effect on respondents. Black and white interviewers differ not only in color but in many other ways, including possibly subtle forms of behavior (for example, the tone of voice or facial expression accompanying a question or in reaction to an answer). Thus, it is only an assumption in any particular study that perceptions by respondents of the interviewer's "race" are what influence answers. Groves focuses on the fact that the white interviewers in the 1968 study were young graduate students and the black interviewers were older professional interviewers, but even if this confound was not important, as Schuman and Converse (1971) argue and as later research by others implies (e.g., Davis 1997), the holistic nature of race-of-interviewer as a variable remains a problem.

The same problem occurs when any other physical attribute is varied (e.g., interviewer gender, as in Kane and Macaulay 1993).[28] The power of experimentation—the nature of internal validity as discussed earlier—comes from the extent to which a single variable is manipulated, with all other features held constant where possible, or where not possible, subject to random variation. But "race of interviewer" and "gender of interviewer" are so complex that only in a large and uncertain sense is there an "it" to be manipulated.

The Three Pens Experiment

I carried out a different kind of survey-based experiment on interviewer effects using a design where the single feature manipulated was simpler than a general physical characteristic of interviewers. The occasion was the highly contentious 1990 election in Nicaragua, where the two main candidates were Daniel Ortega, representing the Sandinista (FSLN) party, and Violeta Chamorro, representing the opposing UNO coalition.[29] As a member of a commission that visited Nicaragua in order to examine the quality of the pre-election polling then under way, I heard repeated concerns about the integrity of the polls. Each of the main political parties reported results over the course of the campaign that strongly predicted victory by its side, which cast doubt on whether the polls were being carried out and reported objectively. Yet some of these same surveys (e.g., one for the *Washington Post–ABC News*) were done by reputable firms and appeared to our commission to be administered carefully. Hence one possibility was that respondents were influenced by their *beliefs* about the political sponsorship of a poll. Furthermore, UNO partisans and also some outside observers claimed that many Nicaraguans felt intimidated by the Sandinista government and would be unwilling to express their true voting intentions to interviewers thought to be tied to the government. Indeed, even interviewers who claimed to be doing a nonpartisan poll might be assumed to be connected to the government unless they gave evidence that this was not the case.

To test these claims, I arranged to have what came to be known as the "three pens" experiment administered approximately two weeks before the election, the cut-off point after which polling was not allowed.[30] Each interviewer carried out her interviews in sets of three, administering them identically so far as practical, except for randomly rotating the particular pen she used to record responses. During one interview the interviewer used a pen that displayed the words "DANIEL PRESIDENTE" and was colored red

and black, the colors of the Sandinista Party. During another interview the pen displayed the letters UNO and was colored white and blue, the colors of the UNO opposition. During the third interview, intended to be neutral, the pen had no lettering and was red and white, colors that did not have a partisan implication. Interviewers introduced the poll as a college class project and did not mention the pens explicitly, nor make direct reference to personal party affiliation or beliefs, so it was up to respondents to draw whatever conclusions they wished from seeing the pens. The experimental design made for a much simpler manipulation of the hypothesized causal variable—a partisan connection to the survey—than is the case when an interviewer's physical characteristic such as race or gender is varied. In addition, by having each interviewer rotate across the three pen conditions, we addressed the point raised by Groves about confounding type of interviewer with overall interviewer variance.[31]

The results for the three pens experiment were consistent with predictions based on the assumption that respondents were influenced by their perceptions of the sponsorship of the poll, as shown in Table 4.8.[32] First, each of the pen conditions identified with a party produced results in line with expectations: the UNO pen condition showed a majority of voters favoring the UNO coalition, and the Ortega pen condition showed a majority favoring the Sandinista party. Second and crucially, the neutral pen condition showed results that also favored the Sandinista party, which fit the assumption that what was intended as a neutral poll was in fact thought by respondents to be connected to the governing Sandinista party. Third, the UNO pen condition turned out to predict accurately the victory by UNO, indeed almost to the percentage point, whereas the other two pen conditions were wrong about the outcome of the election.

Table 4.8 Results of the three pen experiment

	Pen condition		
Vote preference	UNO	Neutral	Sandinista
Chamorro	56%	40%	37%
Ortega	44	60	63
Total	100	100	100
N	(48)	(48)	(57)

Gamma = .26, -SE = .13, $p < .025$ (one-tailed, as predicted)

Source: The original data on which this table is based appear to have been lost in the course of moves by each of the two authors; percentages are taken from Table 3 of their full report (Bischoping and Schuman 1992).

The election itself was carefully monitored by outside organizations such as the United Nations and the Carter Commission, the voting was secret, and there was a consensus among neutral observers that the election was not only fair but was widely believed to be fair. Of course, we cannot be entirely sure that most voters did believe this to be the case, but the successful prediction of the results for the two pens identified with the main parties are clear in any case, and the inference about the result for the neutral pen is quite plausible. The design and outcome with regard to the survey-based experiment itself are strong, even if the associations with the final vote tabulation must be regarded as less certain.

Respondents in each survey condition can be thought to have been influenced by their beliefs about the partisanship of the poll they experienced, and in the case of the UNO pen condition this influence appears to have freed some respondents from apparent intimidation by Sandinista influence and encouraged them to state their actual pro-UNO preference. Paradoxically, bias on the part of a survey organization by indicating its sponsorship may have been necessary to reduce bias on the part of respondents.

The three pens experiment allowed us to study the meaning of responses to polls in a society where the honesty of the polls themselves was not trusted. More generally, polls that attempt to predict a final vote require not so much true attitudes—which are never easy to know for sure—but consistency in the way attitudes are expressed from one setting to another. One might obtain good predictions even in a society where *both* the polls and the election are widely assumed to be rigged—so long as the rigging was in the same direction.

Despite the statistical significance of the results and their plausibility, this investigation is one for which construct replication of some type would be highly desirable. Both the hypotheses and results were unusual, the total sample small because of time and funding constraints, and the supervision necessarily indirect. The design does, however, suggest what might be done in studying elections where there are uncertainties about the pre-election polls, the actual voting, or both.

Conclusions about Context and Interviewer Effects on Responses

Standing back from both questionnaire context effects and interviewer effects, we can see that each of the steps in the survey method introduces the possibility of one or more artifacts. Most of these dangers, however, allow opportunities for deeper understanding once we view them as facts of life. A naïve approach to polls that ignores the context in which the question-answer process takes place is quite likely to trip and fall right on its artifact. But a more disciplined and imaginative approach can treat such effects as an important part of the data, and as an invitation to apply survey-based experimentation, supplemented by interpretive Why questions, to discover larger meaning. "Errors," Isaac Newton wrote in 1686, "are not in the art but in the artificers."

— 5 —

The Survey World and Other Worlds

The eminent psychologist Edward Tolman once reflected on why he had spent so much of his life studying the behavior of rats. Noting their many virtues and few vices, he concluded enthusiastically that rats "are marvelous, pure, and delightful" (1945, p. 166). So also the r(ow) by c(olumn) table to the survey analyst—especially if one or two more variables can be added to enrich the tabulation.

Yet we need to avoid allowing a single set of survey data become our entire way of approaching the social world. It can seem too much a parody of William Blake's "Auguries of Innocence": "To see a world in a grain of sand . . ." This chapter attempts to take note of the connections between the world of a single survey and the evidence available from other worlds. I start at a modest level with the possibility of comparing survey data from one population with survey data from another quite different population, then proceed to consider the linkage of surveys to other methods for studying the social world, and finally end by considering what has come to be known as the attitude-behavior problem. Throughout the chapter I try to keep in mind my broader concern with both method and meaning.

Comparisons across Differently Conceived Populations

Reports of most surveys are based on analysis of responses gathered from a single sample or from repeated samples from the same population over time. Often, however, the overall results of a survey can be illuminated by comparing them with findings based on a survey of a theoretically distinct population. Stouffer's [1955] 1992 book, *Communism, Conformity, and Civil Liberties,* discussed earlier in Chapter 2, provides a classic example. In

addition to his innovative design and analysis of a national cross-section sample of nearly five thousand American adults, Stouffer drew a sample of 14 types of "community leaders" (mayors, Republican and Democratic county chairmen, American Legion commanders, etc.) from 123 mid-sized Americans cities—1,533 individuals in all. He apparently was responding to criticisms from Herbert Blumer (1948) and others that surveys are too atomistic in their approach, failing to take account of the structure of public opinion and the influence of individuals in leadership positions at all levels.

Among other findings from his comparison of the two theoretically distinct populations, Stouffer reported that "Without exception, each of the 14 types of community leaders tends to be more willing to respect the civil rights of Socialists, atheists, . . . and self-avowed Communists than . . . the rank and file in the same cities as the leaders" (Stouffer [1955] 1992, p. 57). He then goes on to argue:

> the fact that responsible community leaders are more likely than the rank and file to give the sober second thought to the civil rights of the nonconformists . . . can be of much significance to America's future. If the reverse had been found, the future might look dark indeed to those who view with anxiety current threats to historic liberties. Plans of public education which aim at building more respect for the American tradition should be able to count on strong support from influential civic leadership at the grass roots. (p. 57)

Thus even without an internal analysis of either sample, which does occur in considerable detail elsewhere in the book, Stouffer was able to argue that his data address a larger issue ignored in most surveys: community leaders stand intermediate between the general public and those in power in Washington, and were less likely to be swayed by attacks on civil liberties by Joe McCarthy than was the broader public. And indeed in the case of the anti-Communist frenzy in Washington in the early 1950s, it is interesting that by the time Stouffer's book appeared, the assault on civil liberties inspired by McCarthy had peaked and was losing its grip on the public imagination.[1]

My own analysis of explanations for why the Vietnam War was a mistake, discussed in Chapter 3, gained greatly from linking a general population survey to a second sample of responses from local college students. Although the student sample was not ideal in the sense of providing a probability sample of a clearly defined population, it was nevertheless reasonable

to claim that it represented a set of individuals influenced by the arguments of those on major campuses who led resistance to the war. The concerns of the student sample turned out to contrast vividly with the considerations that turned the larger public against the war, and the comparison made it clear how different the latter were from the emphases of the anti-Vietnam protest movement on major campuses. Moreover, the contrast continued after the Vietnam War to have implications for different reactions from different parts of the population to U.S. interventions in later years, probably including the invasion of Iraq.

A second example that shows the value of a cross-population comparison started from the opposite direction: not from a general population sample but from a focus on a small set of politically extreme Americans in the Metropolitan Detroit area. David Duke, a former American Nazi and Ku Klux Klan activist, ran for governor of Louisiana in 1991, and as an articulate and nationally known critic of affirmative action and other racial issues, he solicited and received financial support from a substantial number of individuals from outside his state. I obtained from Louisiana the official list of all contributors to the Duke campaign from Metropolitan Detroit—thus creating an initial sample of "Duke contributors."[2] Together with colleagues, I then located and visited the streets on which the contributors lived, and we created a second sample of their near neighbors (usually three neighbors from each street). We next were able to draw a cross-section sample of white households representative of the entire Metropolitan Detroit area. All three samples (Duke contributors, their neighbors, and the cross-section) were sent the same brief mail questionnaire, and the results led to a number of interesting conclusions:

- Duke contributors were spread around the entire metropolitan area, rather than being clustered in any single neighborhood, and in no case were there two or more contributors living on the same street. Thus they were isolates in terms of residence, connected only by their common affinity for what Duke represented.
- Contrary to our expectation, Duke contributors were not more frequently southern-born than the other two samples, though they were more often male, older, and better educated. They also answered our mail questionnaire at a very high response rate (an astonishing 86 percent, as against 78 percent and 74 percent, respectively, for their neighbors and the cross-section sample), and this despite written comments

suggesting that the Duke contributors saw our survey out of the University of Michigan as quite likely biased in a liberal direction.[3]

- In terms of response content, Duke contributors were more apt to perceive crime in their neighborhoods to be a big problem than were either their neighbors or other whites in the Metropolitan cross-section. They also showed less trust in American institutions such as the presidency, courts, media, and unions and expressed more conservative attitudes generally (e.g., opposition to gun control, support for allowing prayer in schools), mostly in a libertarian direction.
- Although Duke contributors differed from their neighbors in the important ways just indicated, their neighbors did not differ significantly from the larger white Metropolitan population in any way we could discover.

Overall and quite remarkably, a survey investigator could evidently obtain what appeared to be an essentially random sample of the white Metropolitan population by first identifying those outliers who contributed money to an ideologically extreme politician running for governor of a different state over a thousand miles away, and then choosing sets of their near-neighbors to include in a survey. Moreover, these comparisons across three theoretically distinct populations could not have been duplicated by sampling a single population, no matter how large the sample obtained.

Survey Results and Results Based on Other Methods

One of the most difficult challenges social researchers face is to gather data using two or more entirely different empirical methods and then make sense of the findings. Of course, in principle different methods aimed at approaching the same problem should lead to the same or at least reconcilable results, but that is not always the case. In addition, measures developed using different methods may be so distinct that a direct comparison of their findings is impossible. It can sometimes seem like results in physics that show light to have properties of either waves or particles, depending on the kind of experiment carried out.[4]

Yet at times a difference in findings can itself prove illuminating, as was the case reported by Converse, Clausen, and Miller (1965) in their study of the Johnson-Goldwater election in 1964. The authors show that poll results

were clear early in 1964 that Goldwater was highly unlikely to win the November election. Next they asked how Goldwater and his backers came to see a quite different reality and pointed to the letters that some of these same survey respondents reported having written to newspapers and public officials during that period of time. If we are willing to assume that answers to factual questions about letter writing are close to what an actual enumeration of such letters would yield, Goldwater held a visible lead in this form of "public opinion"—opinion as reflected in the letters. Thus the two different sources of information about reality or, indeed, the two different realities helped explain how Goldwater could be nominated with high hopes for his election and yet go on to lose by a landslide.

An unusual study by Nisbett and Cohen (1996) drew on a number of different research methods to test a proposition about the location in the southern United States of a "culture of honor." With both secondary analysis of existing survey data and new survey data of their own, they confirm that southern white males are more likely than northern white males to endorse the use of violence when situations are described that involve an insult to oneself. They then support and explicate these findings with additional evidence based on other methods, for example:

- Department of Justice and other data show homicides to be more common for whites in the South than in the North, and specifically in rural southern areas that were settled originally by peoples from the fringes of Britain who had been part of a herding economy and who continued in that tradition in the United States.
- To show that actual behaviors, as well as verbally expressed attitudes, reflect the overall regional difference, actions by university male students from the South are compared to those of male students from the North in an experimental situation that entailed a personal insult. The measured responses included observable facial expressions, completions of stories involving insults, and two relevant physiological indicators (cortisol and testosterone). Almost all of the predictions of South-North differences among students were supported.
- The authors report that both laws and votes of political representatives from southern states are less strict with regard to guns and give more support to violent action in self-defense, provide more backing in

Congress for military actions, are more lenient toward domestic violence, and allow more executions of convicted prisoners. These and other measures are used to point to southern distinctiveness at the level of institutions and institutional representatives.

In short, the Nisbett and Cohen (1996) book, *Culture of Honor: The Psychology of Violence in the South,* is a tour de force from a methodological standpoint, with survey data only one element provided in support of an overall argument.[5]

My own attempt at combining survey data with other types of evidence started from a major effort to use polls to describe and understand changes in racial attitudes between 1942 (the first available measures in national surveys) and the late 1990s (Schuman et al. 1997). On basic issues of racial discrimination ranging from public areas such as employment to more personal spheres such as intermarriage, the trends in surveys over the past several decades pointed clearly toward increased acceptance by white Americans of the principle of equal treatment. Figure 5.1 illustrates the change, using two different questions about racial intermarriage, one having to do with public laws and the other with personal preferences.

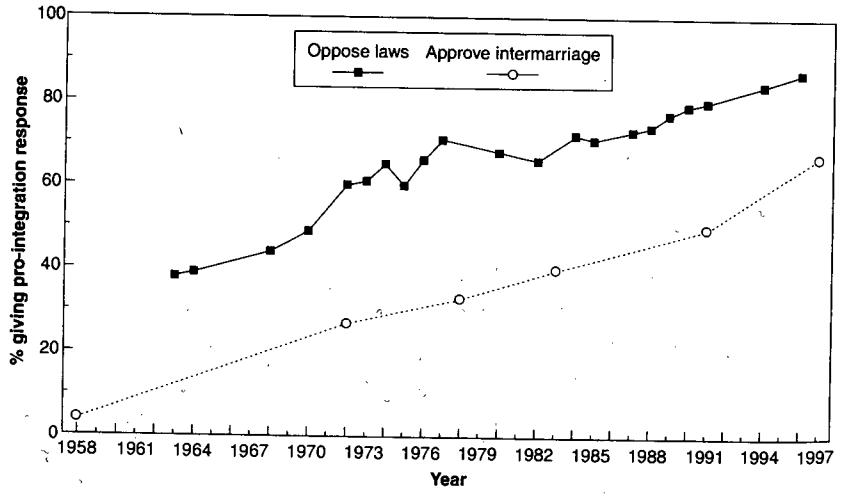

Figure 5.1 Comparison of laws against intermarriage and approval of intermarriage. *Source:* Schuman et al. (1997, p. 118). Copyright © 1985, 1987 by the President and Fellows of Harvard College. All rights reserved.

Although at each time point there was greater support for eliminating laws against intermarriage than there was personal approval of intermarriage, responses to both questions showed essentially the same trends over time (Schuman et al. 1997).[6]

As part of my overall program of research, I was able to obtain entirely different types of data of two kinds. First, Ronald Humphrey and I coded the presence of African Americans in magazine advertisements in two magazines: *Time* and *Ladies Home Journal* from the 1950s into the 1980s (Humphrey and Schuman 1984). During that period blacks were increasingly included in magazine ads and increasingly shown in equal status work situations. However, consistent with the survey data, scenes of blacks and whites in social situations remained relatively rare, and there were other subtle respects in which signs of inequality remained (e.g., white authorities were often shown supervising blacks, but not the reverse).

Second, a much more difficult companion study was an attempt to code actual behavior toward black couples in comparison with behavior toward white couples in a sample of New York City restaurants in 1981 (Schuman et al. 1983) and then to compare what we found at that point in time with the results from a previous investigation three decades earlier (Selltiz 1955). Carrying out a systematic study of social interaction of this type proved to be extremely difficult and was hampered by problems such as the ambiguity of treatment that black couples reported experiencing (e.g., "slow service")—a problem akin to, but more difficult than, coding complex open responses. We also encountered serious racial tensions between the white and black individuals whom we recruited to serve as "testers" of treatment by restaurant personnel. In addition, the study was limited in size because of difficult logistics and minimal funding, though in the end we did find some evidence of a decrease in discrimination between the time points of 1950 and 1981.The research in this case was less instructive for its substantive conclusions than for its lessons in the challenges of systematic social observation of natural behavior that can parallel responses to surveys and polls.[7]

A final example to consider is not a combination of methods already achieved but a set of survey results that demands additional nonsurvey evidence. In a national sample in 1978, Presser and I assessed attitudes toward requiring permits for guns—an issue thought to have influenced some past

elections because of the strong feelings of those opposed to any steps toward regulating firearms. One key question was the following:

> Would you favor a law which would require a person to obtain a police permit before he could buy a gun, or do you think such a law would interfere too much with the right of citizens to own guns?[8]

Favor	61%
Oppose	39
	100
N	(1,076)

Although the majority of Americans clearly favored requiring gun permits in answer to this question, our main interest was in how supporters and opponents answered two follow-up inquiries, shown in Table 5.1. "Centrality" was intended to provide a subjective measure of attitude strength; "committed action" was included to obtain a self-report of actual behavior. Pro- and anti-permit respondents do not differ greatly in terms of the centrality of this issue, but they do show a large and highly significant difference in their actions: 20 percent of those opposed to gun permits report having written letters, given money, or done both, whereas only 7 percent of those favoring permits say they have done the same—a nearly 3 to 1 ratio. Equally striking is the clear association of attitude centrality to committed action for opponents of gun permits, as shown in Figure 5.2, whereas there is no sign of such an association for those favoring permits.

Thus opponents of gun permits take political action consistent with the degree of strength of their attitudes, whereas those who support requiring permits not only take many fewer actions but the actions they do take show no relation to their attitudes. How can we account for this sharp difference? It seems unlikely that it can be attributed entirely to individual differences in attitudes, since that would not account for the difference in relationships shown in Figure 5.2. What seems more likely, especially based on many reports about the size and efficiency of the primary lobby opposed to gun control—the National Rifle Association (NRA)—is that it has developed highly effective methods of targeting its most committed members at strategic points during elections and legislative processes in order to stimulate letters and contributions. Organizations favoring gun permits and other gun control legislation may be not only smaller and weaker but also less successful at mobilizing those who agree strongly with them about gun

Table 5.1 Centrality and committed action on gun permits

Centrality: "How important is a candidate's position on permits for guns when you decide how to vote in a congressional election—is it one of the most important factors you would consider, a very important factor, somewhat important, or not too important?"

Committed Action: "Have you ever written a letter to a public official expressing your views on gun permits or given money to an organization concerned with this issue?"

	Favor	Opposed
Centrality		
One of most important	5%	8%
Very important	22	26
Somewhat important	39	28
Not important	34	38
Total	100	100
N	(313)	(208)
	$\chi^2 = 6.3$, df = 3, p = .10; $tau_b = .00$	
Committed Action		
Written letter	4%	6%
Given money	2	8
Both letter and money	2	6
Neither	93	80
Total	100	100
N	(653)	(417)
	$\chi^2 = 46.4$, df = 3, p < .001	

Source: University of Michigan Survey of Consumer Attitudes, August 1978.

control issues; hence even decisions to send letters and funds remain more with individuals, rather than elicited as part of well-run campaigns.

There is no single way to test the hypothesis just advanced, but one type of useful information could come from a content analysis of pro– and anti–gun control publications, in order to determine when they identify an issue or candidate as calling for action and what steps each takes to mobilize members. It would also be of value to track legislation likely to prompt such efforts. In addition, if respondent cooperation can be obtained, perhaps by using incentives, a panel of individuals could be asked to save dated messages they receive about gun control issues. Letters to the editor in local and

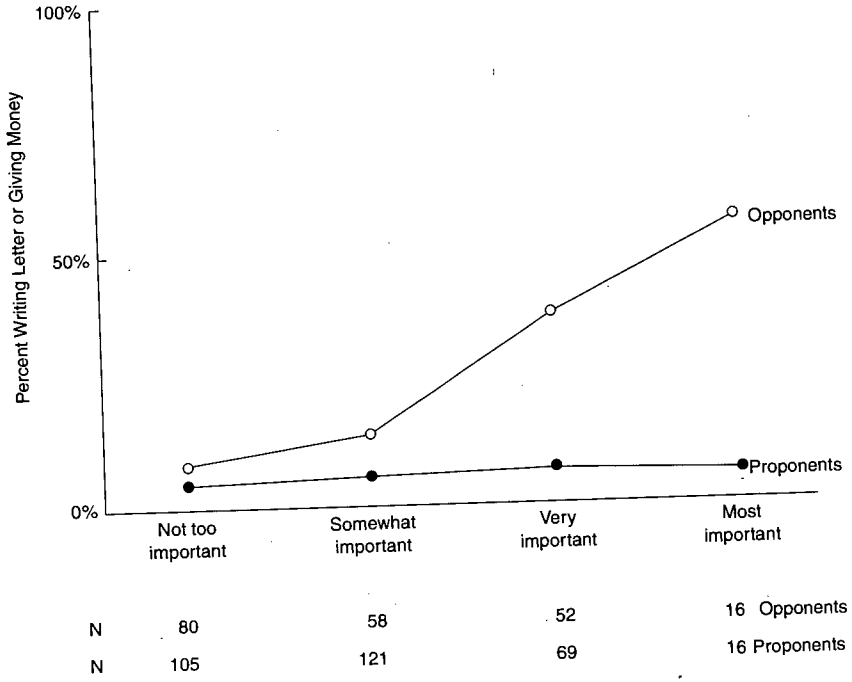

Figure 5.2 The relation of behavior to centrality for gun-permit proponents and opponents. The behavioral measure (the vertical axis) is dichotomized into those showing any behavior (letters, money, or both) and those showing none.
Source: Schuman and Presser ([1981] 1996, p. 242)

national publications could also be reviewed for evidence about the side taken, the specific content, and especially the timing. Once a decision is made to supplement survey data with other evidence, it is likely that other creative steps could be developed to link attitudes to organizational effectiveness, with the goal of explaining the large difference in committed action by those for and against increased control of guns. This study provides a good example of why we need more effort directed at connecting the survey world with evidence from other worlds.

The Relation of Attitudes to Behavior

Concern about different methods of research points to an issue important to almost all attitude surveys: the extent to which attitudes expressed in response

to a questionnaire are consistent with directly relevant behavior outside the survey. The issue was raised initially by L. L. Thurstone (1931), the psychologist most responsible for introducing the systematic measurement of attitudes into social science. But the attitude-behavior problem was then given much greater force as the result of a remarkable article by Richard LaPiere (1934), which argued for an almost ineluctable gulf between attitudes as measured by questionnaires and actions as observed in real situations.

LaPiere began with a picturesque thought experiment about a question that he conjectured, perhaps whimsically, might have been put to people in the early 1930s: "Would you get up to give an Armenian woman your seat in a street car?" No matter how carefully the question was constructed and posed, LaPiere claimed that it could obtain only a symbolic response to a symbolic situation:

> The words 'Armenian woman' do not constitute an Armenian woman of flesh and blood, who might be tall or squat, young or old, well or poorly dressed . . . And the questionnaire response, whether it be yes or no, is but a verbal reaction and this does not involve rising from the seat or stolidly avoiding the hurt eyes of the hypothetical woman and the derogatory stares of other street-car occupants. (p. 230)

Having made his primary theoretical point, LaPiere went on to present data from his travels across the country with a foreign-born Chinese couple in the early 1930s. It was a time when discrimination against Chinese, along with many other minorities, was blatant and largely unquestioned in the United States. Yet LaPiere found that his Chinese companions were accommodated without difficulty in all but one of the 250 restaurants, hotels, and tourist homes at which they stopped. Then, six months later, after the completion of his travels, he sent a questionnaire to each place they had visited and asked if it would "accept members of the Chinese race as guests." Of the 128 responses he received, all but one replied No or indicated Uncertainty. (In a useful addition, he also sent questionnaires to a sample of hotels and restaurants that were in the same regions but that they had not visited, and obtained similar results.) LaPiere then concluded that questionnaires, though easy to administer and able to produce quantitative data, yield results that are of no relevance if the goal is to understand actions that occur in real situations. He also notes that the factors that shaped acceptance of his Chinese companions had mostly to do with their clothing, their cleanliness and neatness, and their baggage, and also with their self-confident smiles as they entered an establishment.

Later reviews of LaPiere's 1934 article have pointed out flaws in the design of his Chinese visitors study (e.g., see Ajzen et al. 1970 and Dillehay 1973), but the evidence of almost total inconsistency between attitudes and behavior in LaPiere's report remains a challenge to those of us who expect attitudes measured in surveys to be related to, indeed to predict, later behavior. Moreover, although LaPiere's own theoretical perspective was entirely from the standpoint of a qualitative sociologist, his assumption about the power of immediate situations is consistent with conclusions in contemporary social psychology that have been drawn from a range of experiments, represented by well-known names such as Asch, Latané and Darley, Milgram, Nisbett and Ross, and Zimbardo.

There are several ways to attempt to meet the challenge set forth by LaPiere and others who share his doubts about the value of survey measures of attitudes for learning about nonsurvey behavior. First, even LaPiere acknowledged that polls can predict well how people will vote in the aggregate, and the accuracy of poll predictions close to an election continues to be documented—and this in the face of increasing difficulty in obtaining adequate samples due to new obstacles to contacting respondents and persuading them to take part in surveys (Traugott 2005). LaPiere treated both pre-election polls and actual voting as basically verbal responses to symbolic situations, but his acceptance of such predictions as valid indicates that even for him, the distance between survey responses and at least one kind of nonsurvey behavior is bridgeable.

Within the survey context itself, we can also do more to encourage people to indicate how they are likely to respond to pressures that we and they know can occur in real situations. For example, in a cross-section sample survey of the white population of Metropolitan Detroit at the end of the 1960s, I described the following situation involving racial discrimination, and asked three different follow-up closed questions in sequence (Schuman 1972):

> Suppose a good black engineer applied for a job as an engineering executive. The personnel director explained to him: "Personally I'd never given your race a thought, but the two men you would have to work with most closely—the plant manager and the chief engineer—both have strong feelings about blacks. I *can* offer you a job as a regular engineer, but *not* at the executive level, because any serious friction at the top could ruin the organization."[9]

A. Was it all right for the personnel director in this case to refuse to hire the black engineer as an executive in order to avoid friction with the other employees?
B. Should the personnel manager have asked the other men how they would feel about working with a black engineer and then made his decision on the basis of their wishes?
C. In general, do you think employers should hire men for top management without paying any attention to whether they are white or black?

In principle, these three questions should have been answered consistently in terms of either accepting or rejecting discrimination, but instead there was wide variation in responses across the questions:

C. Agrees to racial discrimination in principle.	13% of the sample
A. Agrees to discrimination for sake of harmony.	41% of the sample
B. Agrees to discrimination if majority favor it.	52% of the sample

Depending on the question asked, one could argue that employment discrimination at that point in time was supported by only a small minority of the white population, by two-fifths of that population, or by just over half. This kind of difference, if demonstrated between survey and behavioral data, could readily be interpreted as strong evidence for attitude-behavior inconsistency. Yet these are differences among responses to questions asked not only in the same survey but almost in the same breath! Respondents showed themselves quite capable of taking a different position on the same basic action once the situation itself was portrayed differently, and they did so, according to our interviewers, with little or no sense of embarrassment. Therefore, without claiming that the entire power of real situations can be captured simply by asking the right set of questions, some of the important social pressures that operate in life can be approximated by constructing questions that convey realistic situational forces.[10]

Beyond the conceptual points already made, there is now considerable evidence for what is required to increase consistency between attitude questions asked in a survey and behavior outside the survey. A fundamental starting point is the need to distinguish between "literal consistency" (do people do what they say they will do?) and "evaluative consistency" (are people ordered in the same way on both attitude and behavioral measures?).[11] Literal consistency requires that the survey attempt to measure

not an attitude toward an object (e.g., a person or a race or a country), but rather an attitude toward a particular behavior toward the object, and to specify as clearly as possible the time and other contextual factors likely to influence the behavior. This is an important correction to earlier research, which assumed that a broad positive or negative attitude toward an individual or a group implied whatever particular action an investigator happened to measure. The assumption can be seen to be naïve once its implications are thought through, for example, giving a particular gift to one's spouse may show love, but from the knowledge that someone loves their spouse, one could not infer that they will give a particular gift (such as flowers, a puppy, or a vacation in Nepal), for they may express their love in other ways.

Evaluative consistency does assume that an attitude toward an object should be consistent with the overall favorability or unfavorability of actions taken toward the object, even though not necessarily with any particular action. Thus it leads to research that measures a set of behaviors and predicts that a score based on the set as a whole will correlate well with the attitude measure, though recognizing that no specific behavior may be predictable. A fine example of exploring evaluative consistency using a natural population was reported by Weigel and Newman (1976) with a field study in one town that showed a measure of pro-environmental attitudes to relate well to a multi-item measure of pro-environmental behaviors (e.g., actual participation in a roadside litter pickup program). The attitude measure correlated, on average, only .29 with individual environmental behaviors, but the association rose to .62 with an aggregated index of 14 different behaviors. Moreover, in considering the size of these relations, it is important to compare them not with a theoretical maximum of 1.0, but with associations we take seriously in other areas of research. For example, we should note that no more than moderate-sized correlations have been found for the relationship between a father's occupational status and a son's occupational status in America (Blau and Duncan 1967), for the relationship between academic aptitude scores and college grades (Cronbach 1970), and for most other nonartifactual associations of interest to social scientists.

An Attitude-Behavior Field Study

Much of the research examining literal consistency has been done with college students and with measures atypical of surveys, but Brannon and I and a group of graduate students (1973) tested literal consistency between a

Table 5.2 Percent of white respondents supporting either of two laws

"Suppose there is a community-wide vote on the general housing issue. There are two possible laws to vote on. [Present card and read:] Which law would you vote for?"	
1. "One law says that a homeowner can decide for himself who to sell his house to, even if he prefers not to sell to blacks."	82%
2. "The second law says that a homeowner cannot refuse to sell to someone because of their race or color."	16
DK, Neither, N.A.	2
Total	100
N	(640)

Source: 1969 Detroit Area Study (Interuniversity Consortium for Political and Social Research).

survey-based attitude measure concerning open housing laws and willingness at a later point to sign a real petition either for or against an open housing law. The research involved a field experiment attached to a standard face-to-face sample survey of the Detroit Metropolitan Area in 1969.[12] White respondents were asked the survey question shown in Table 5.2, and then three months after the survey had been completed, a different person representing a group of "concerned citizens" came to their door and asked the respondent to sign a petition to be sent to the governor and other state officials. The two petitions, only one of which was presented to a respondent, read:[13]

> [*Owner's Rights Petition:*] We believe that it is the right of each homeowner to decide for himself to whom he will sell his house. We urge you NOT to support any law which would force homeowners to sell to someone against their wishes, and if such a law is passed by the current legislature we ask you to veto it.
>
> [*Open Housing Petition:*] We believe that a person who has a home for sale Does Not and SHOULD NOT have the right to choose buyers on the basis of race, color, or religion. We therefore urge you to support legislation this year which will end racial discrimination in housing, once and for all, in Michigan.

Those who agreed to sign the petition were then asked to perform a still stronger action by allowing their name to be included if the petition was published in one of the two major Detroit newspapers.

More than 80 percent of the respondents in the survey had indicated opposition to an open housing law. These individuals (N = 525) were divided randomly into a three-quarters subsample asked to sign the Owner's Rights petition (Condition A), and a one-quarter subsample asked to sign the Open Housing petition (Condition B). Most of the remaining respondents from the original interview had indicated support for an open housing law, and all these people (N = 101) were presented the Open Housing petition (Condition C) because the subsample seemed too small to break further. Ideally we would have created a similar random allocation for those favoring open housing in the survey, but we thought it better to maintain the group intact for more reliable estimates of their action.[14]

The results of the experiment are shown in Table 5.3. Over two-thirds of the respondents on both sides of the issue (Conditions A and C) were willing to affirm their survey-elicited stand by signing a petition ($\chi^2 = 9.44$, $df = 1$, $p < .01$). The percentages declined, though not remarkably, in both conditions when stronger action was requested: nearly three-fifths signed with publicity in both conditions. Those espousing owner's rights in the survey showed somewhat more willingness to sign an Owner's Rights petition (85 percent) than did open housing proponents to sign an Open Housing petition (70 percent), but of those who agreed to sign a petition, the open housing proponents were more apt to follow through with stronger action, and thus the two sides did not differ at all if signing with publicity is considered the criterion of action.

Although a high proportion of the sample was consistent in behaving in accord with their survey response, such apparent affirmation could be exaggerated by the social pressure to sign a petition at all. In Condition B we see that 78 percent of those who had favored owner's rights in the survey were also consistent in refusing to sign an Open Housing petition, though 22 percent did sign that petition opposed to the position they had supported in the prior interview (assuming that reflected their "true attitude"). Thus most respondents did not simply sign whatever was presented to them. Moreover, declining to sign with publicity is much greater as a proportion of overall signing in Condition B than in Conditions A or C, indicating the limit to situational forces in this case.

To estimate overall consistency for owner's rights respondents, one can average those in Condition A who acted consistently and those in Condition B who were consistent in withholding action. If the two conditions are weighted equally, then of those who had originally favored owner's rights,

Table 5.3 Consistency between attitude question in survey and later behavior

Signing conditions	Attitude preference	Housing petition presented	Percentage taking each action			Total	N
			Refused to sign petition	Signed but declined publicity	Signed and agreed to publicity		
Condition A	Owner's rights	Owner's rights	15%	26	59	100	293
Condition B	Owner's rights	Open housing	78%	10	12	100	85
Condition C	Open housing	Open housing	30%	12	58	100	76

Source: Survey data and behavioral data are both from the 1968 Detroit Area Study and are available in the same file at the Interuniversity Consortium for Political and Social Research (ICPSR).

85 percent were consistent in signing the Owner's Rights petition, and 78 percent were consistent in refusing to sign the Open Housing petition, which gives an overall average of 82 percent consistency.[15] This average, together with the consistency results for those who supported open housing in the survey, indicates that so far as we can judge from this single field experiment, attitudes can predict behavior to a reasonable extent, though of course not perfectly.

In one sense, of course, LaPiere might have claimed that the attitude-behavior experiment reported here had a large symbolic component, because both in the survey and in the action phase the descriptions of laws were presented in the form of words, though the latter involved presentation by real individuals and carried the implication of wider public knowledge. Before seeing this verbal character of the action as a serious limitation, we should recognize that many highly important actions occur in the form of words stated or written ("I Do" in marriage ceremonies, signatures on checks and contracts and in military enlistments, acts of congress and presidential vetoes), and such actions can clearly entail consequences of a nonsymbolic nature (e.g., prison for failure to live up to military enlistment). Thus the line between symbolic and nonsymbolic actions is not as clear-cut as LaPiere implied.

Yet are there situations for which no survey is likely to be able to predict behavior? Certainly at the extreme this seems likely. How would you behave if a flying saucer suddenly circled around your house? That is probably not predictable, nor are many other unusual events. The main point here can be brought down to earth in the form of a generalization by Fazio (1986, p. 219): when an attitude "is grounded in and based on prior behavior, the attitude-to-later-behavior relation is stronger than when the attitude is based on indirect experience." Fazio presents evidence that supports his finding, which indeed would be supported by our intuition.[16]

Some Concluding Thoughts on Studying More than One World

Asking the same questions to quite different theoretical populations can often be useful, as it was in Stouffer's 1954 survey of the general public and of community leaders. Even without a clear hypothesis, this kind of comparison helps investigators to understand the scope of the findings discovered in their primary sample. Another version of such an extension is replication of the same basic relation within different countries to discover its generality.

For example, in exploring the proposition that national and world events that people recall as important are those that happened during their adolescent and early adult years (discussed in Chapter 2), it has been useful to show that the same basic relation for social memory could be found in samples in both Germany and Japan, with both the same event (World War II) and different events (reunification for Germans, death of the emperor for Japanese)[17].

Use of radically different methods to pursue the same basic investigation is much more challenging, though multimethod studies are now becoming a frequent and admired approach to research. The question arises as to what an investigator should expect in the way of agreement when the methods produce entirely different types of data. In my case, I showed that there is evidence of less racial discrimination today than in the past in survey responses, in magazine advertisements, and in the treatment of customers in restaurants, but beyond that general finding it is difficult to construct comparable variables that can be examined with much precision in three very different sets of data. Nisbett and Wilson argued from federal statistics on violence that the most crucial part of Southern heritage involves rural areas settled originally by peoples from the fringes of Britain who had been part of a herding economy; however, their laboratory experiments on actual behavior compared general samples of students from the "South" and the "North" who happened to be present in a single northern university setting, and the authors do not (and perhaps could not) focus on subregions to fit their earlier theorizing. The challenge of different methods often becomes one of comparing apples and oranges, or perhaps apples and string beans—different in a great many ways. So although the goal of approaching a problem with entirely different methods deserves strong encouragement, I have learned that it should not be undertaken lightly and without considering carefully whether and how it will be possible to synthesize the results.

When we turn to the issue of whether attitudes and actions are related, there is the fortunate possibility of profiting from inconsistencies between different forms of data. LaPiere's travels with a Chinese couple may have presented proprietors with situations even more novel than LaPiere himself had in mind. His goal might have become understanding why hotel and restaurant proprietors rejected reservations for Chinese visitors in question form, but accepted LaPiere's Chinese companions in actuality. When answering questions by mail, did proprietors picture Chinese as laborers in pigtails and coolie hats, speaking an unintelligible tongue, and thus not

even recognize the real couple as Chinese? Or did the proprietors act in terms of an overriding belief that the less disturbance the better? LaPiere could have attempted to learn more about these and other possibilities. One good way, not perfect by any means, but among the best available, would have been to ask open-ended follow-up questions, as recommended in Chapter 3, and then to record each proprietor's personal definition of the situation. If he had tried to do this, with the goal of representing a meaningful population of proprietors, of proceeding systematically so as to avoid bias in his inquiry, and of connecting the information to social categories such as age and education, then LaPiere would have reinvented the attitude survey in its richest form. He might have used his negative results not as a stopping point for rejecting the survey approach, but as a starting point for the use of the survey as a method to search for new meaning.

— 6 —

Hunting a Social Science Snark

"Just the place for a Snark," the Bellman cried,
As he landed his crew with care . . .

"Just the place for a Snark! I have said it twice:
That alone should encourage the crew.
Just the place for a Snark! I have said it thrice:
What I tell you three times is true."

. . .

"But oh, beamish nephew, beware of the day,
If your Snark be a Boojum! For then
You will softly and suddenly vanish away,
And never be met with again!"

—LEWIS CARROLL (1874)

"The great tragedy of Science," wrote Thomas Huxley (1870, p. 244) is "the slaying of a beautiful hypothesis by an ugly fact." Tragic perhaps, but essential if progress is to be made in social research. At the same time, it is also true that a highly plausible theoretical conclusion should not be given up lightly in the face of disconfirming data, at least not without careful consideration to make sure the fault does not lie with limitations in the evidence. In a chapter titled "Difficulties of the Theory," Darwin ([1876] 1988, p. 140) wrote: "Some of the [difficulties] are so grave that to this day I can hardly reflect on them without being in some degree staggered . . ." He did his best to address the difficulties, but above all he found his theory of natural selection so persuasive that he could not reject it because of unsolved empirical problems.

No social science theory has anything even faintly resembling the scope and power of the theory of natural selection, but data and theory can at times be competitive in social research in something of the same way, even if on a much smaller scale. The present chapter recounts a case from one part of my earlier research where the competition between meaning and method initially led to compelling data that stimulated what we believed to