

Violating Conversational Conventions Disrupts Cognitive Processing of Attitude Questions

Allyson L. Holbrook and Jon A. Krosnick

The Ohio State University

Richard T. Carson

University of California, San Diego

and

Robert Cameron Mitchell

Clark University

Received April 20, 1999; revised August 16, 1999; accepted August 18, 1999

This research distinguishes conversational norms from conversational conventions and tests the notion that violation of conversational conventions in attitude questions disrupts processing and reduces data quality. Our first study showed that in questions with simple, dichotomous affirmative and negative response alternatives, respondents expect the affirmative response alternative to be offered before the negative one. Four studies showed that

The authors thank Stanley Presser, Raymond Kopp, W. Michael Hanemann, Paul Ruud, V. Kerry Smith, Michael Conaway, Kerry Martin, Richard Bishop, Trudy Cameron, Carol Jones, Norman Meade, Pierre Du Vair, Alan Randall, Penny Visser, Joanne Miller, George Bizer, and Dean Lacy for their help and advice throughout this project. We also thank David Moore and the Gallup Organization for providing access to the data analyzed in Study 6. The work described in this paper was funded in part by National Oceanic and Atmospheric Administration Contract 50-DGNC-1-00007 and by a grant under the Cooperative Research and Training Program sponsored by the Minerals Management Service, U.S. Department of the Interior, and the Coastal Marine Institute, University of California (Task 12390). The research reported here was conducted partly while the second author was a Fellow at the Center for Advanced Study in the Behavioral Sciences (supported by NSF Grant SBR-9022192).

Address correspondence and reprint requests to Allyson L. Holbrook or Jon A. Krosnick, Department of Psychology, The Ohio State University, 1885 Neil Avenue, Columbus, Ohio 43210. E-mail: holbrook.52@osu.edu or krosnick@osu.edu.

violating this convention disrupts cognitive processing. Respondents took longer to answer questions asked unconventionally, and their answers were less predictable by their attitudinal dispositions and their demographic characteristics. In addition, answers were less responsive to manipulation of a key feature of an object of judgment. Another study ruled out a possible alternative explanation for the effect (an increase in disconfirmatory thinking) and documented that convention violation led people to generate more thoughts irrelevant to the question's topic. These disruption effects were reliable only among people for whom the cognitive work entailed by generating optimal answers to questions was most difficult. © 2000 Academic Press

Everyday conversations unfold efficiently because speakers conform to norms and listeners presume they conform to such norms (Clark, 1992, 1996; Grice, 1975; Levinson, 1983). When constructing an utterance, people take into account the perspectives and knowledge of their conversational partners and try to make their comments clear, relevant, honest, and appropriately informative. Because speakers and listeners operate with such principles or norms in mind, it is possible to convey background assumptions, nuances, points of emphasis, hesitations, and more without having to express them explicitly with words. They are suggested instead by the *way* something is said.

Conversational norms apply not only to everyday communications, but also to the communication between questionnaire designers and respondents (for a review, see Schwarz, 1996). Respondents interpret the meaning of questions presuming that the question writers are conforming to conversational norms, and this interpretative approach produced what at first appeared to some scholars to be poor reasoning by respondents in a number of studies. For example, the order in which people provide pieces of information to one another conveys the perceived importance of the pieces of information (Krosnick, Li, & Lehman, 1990). This turned out to have been partly responsible for evidence suggesting that people ignore base rate information in making probability judgments (Kahneman & Tversky, 1973). Likewise, people presume that speakers provide information that they believe to be relevant to the conversation at hand. This was partly responsible for people's apparent underuse of diagnostic information when they were also given nondiagnostic information (Nisbett, Zukier, & Lemley, 1981; Tetlock, Lerner, & Boettger, 1996). In both of these cases and most others in this literature, researchers did not conform to a conversational norm when asking a question, but listeners presumed that they did, thus leading to misunderstandings (for other examples, see Dulany & Hilton, 1991; Schwarz, Strack, Hilton, & Naderer, 1991; Strack & Schwarz, 1992; Strack, Schwarz, & Wänke, 1991).

In this paper, we make a distinction between conversational norms and conversational conventions. Conversational norms are implicit rules and assumptions that are followed and made by speakers and that listeners assume speakers follow. These conversational norms allow speakers to communicate more than simply the content of their words, and when a conversational norm is violated by a speaker, misunderstanding may occur, because listeners generally assume that

norms are followed. In contrast, we define conversational conventions as arbitrary customs that are followed by most speakers but communicate no added meaning. Therefore, violation of a conversational convention is unlikely to change the meaning of what is said, but listeners may notice the violation and perhaps even be distracted by it.

Our vehicle for this exploration is a conversational convention involving the order in which response alternatives to closed-ended questions are provided. We focus on a very common question format, involving simple, dichotomous, mutually exclusive affirmative and negative response alternatives. Such questions are commonly used in laboratory research and in surveys, asking people whether they favor or oppose a government policy, approve or disapprove of the president's job performance, agree or disagree with an assertion, or would vote for or against a proposed new law. Changing the order in which these response alternatives are offered in a question does not change the meaning of the question to the listener. But we suspected that there is a conventional way to offer these response alternatives and that violating this convention compromises the quality of data obtained. We begin below by proposing and justifying our hypotheses, and we then report six studies testing them.

HYPOTHESES

Although some conversational norms are useful because they facilitate communication of unspoken information, conversational conventions do not add to the volume of information transmitted through communication and seem to have evolved simply out of habit. Specifically, much of grammar constitutes shared habitual approaches to combining words into sentences. In English and German, for example, it is customary to present attributive adjectives before the noun to which they refer, as in "I bought a gray dress" (see Garcia Rojas & Molesworth, 1996) or "ein grau kleid" (see Hammer & Durrell, 1991). In English, one would not say "I bought a dress gray."¹ Someone hearing this sentence might still be able to infer its intended meaning, but he or she is likely to be surprised and at least momentarily distracted by the grammatical violation. Yet in French, Spanish, and Italian, this would not be surprising at all, because it is customary to present attributive adjectives *after* the noun to which they refer: "une robe grise" (in French; see Hawkins & Towell, 1996), "un vestido gris" (in Spanish; see Butt, 1996), or "un vistito grigio" (in Italian; see Grandgent & Wilkins, 1915).²

There are other shared customs in the construction of utterances that go beyond grammatical rules as well. For example, restaurant personnel and customers alike conform to a script outlining the order in which particular conver-

¹ We are referring to single-word adjectives that modify a noun. When adjectival phrases are used to modify a noun (e.g., "the dress that is gray"), the adjective can come after the noun.

² Although the *convention* is for adjectives to come after nouns, in each of these three languages, there are exceptional adjectives that typically come before the noun. These exceptions vary somewhat in Spanish, French, and Italian, but typically include adjectives of size (e.g., large, small) and beauty (e.g., beautiful, pretty; see Butt, 1996; Grandgent & Wilkins, 1915; Hawkins & Towell, 1996).

sational events occur (Shank & Abelson, 1977). Servers usually ask first for drink orders, then return later to take orders for appetizers and entrées, and finally return after the entrées are eaten to inquire about any interest in dessert. This order has some basis in practicality, because it is easy to serve drinks quickly (thereby allowing patrons to assuage their thirst), and because people's preferences regarding dessert are often determined by their moods following the main part of a meal.

Consider what might happen, though, if this script is violated. For example, imagine a customer who has been given a menu, has placed a drink order, has received a beverage, and is deep in thought, struggling to choose between two mouthwatering entrées. "The crab cakes look great, but the broiled chicken breast would be healthier," go her thoughts. Just as she is beginning to resolve the conflict by reviewing how healthy her recent meals have been and decide if she is entitled to splurge, her server approaches smiling broadly and says enthusiastically, "What would you like for dessert today?"

In light of the restaurant script, the customer is likely to be surprised by this utterance: "Did the server forget that I haven't yet ordered and eaten my appetizer and entrée? Did the server confuse me with another patron who said she was only having dessert and nothing else? Is this restaurant well known for its soufflés, which must be ordered at the beginning of a meal in order for them to be ready at the appropriate time?" For at least a moment, some such possible explanations may flash through the customer's mind, perhaps distracting her just enough so that she forgets which entrée she was leaning toward and why, or perhaps changing the process by which she determines which entrée she would like (by considering its implications for dessert later).

Linguists have argued that similar nongrammatical conventions govern the order in which words are listed in sentences during everyday conversation (Cooper & Ross, 1975). According to Cooper and Ross (1975), each of the following represents a conventional ordering that sounds natural, and each utterance would instead sound surprisingly unnatural if reversed: "here and there," "this and that," "now and then," "sooner or later," "parent and child," "mother and daughter," "man and wife," "once or twice," "first or second," "cowboys and Indians," "person, place, or thing," "animal, vegetable, or mineral," "land and sea," "field and stream," "front and back," "hunter and hunted," "the living and the dead," "life and death," "at home and abroad," "form and substance," and "abstract and concrete."

Cooper and Ross (1975) also identified another convention of ordering: that when two terms are being listed, one positive and the other negative, it is conventional to list the positive one first, as in "like or dislike," "for or against," "support or oppose," or "friendly or unfriendly." Therefore, when a question is asked to gauge a person's attitude during everyday conversation, it would be conventional for a speaker to offer the positive or affirmative response choice(s) first and the negative response choice(s) second, as in "Are you going to buy it

or not?," "Are you going to vote for it or against it?," and "Do you support or oppose it?"

If this is a conversational convention, then offering the negative alternative first in a questionnaire item with simple, dichotomous positive and negative response alternatives would violate this convention and might therefore be surprising to respondents. At the very least, such a violation is likely to be momentarily distracting, pulling some cognitive attention away from simply answering the question, because one registers (even if unconsciously) the unexpected violation of conventional ordering.

For a respondent who is highly skilled at the cognitive exercises involved, this momentary distraction may have little impact on what follows. But a respondent who has more difficulty performing the mental operations involved in question answering may be particularly affected by the violation and may then find it hard to get back on the cognitive track, so to speak. Consequently, the retrieval, integration, and translation of information required for generating high-quality data (see Tourangeau & Rasinski, 1988) may be executed less well, introducing error into the answers people give. Consistent with this notion, research examining response effects has demonstrated that small changes in question wording are more likely to affect the responses of respondents low in cognitive skills than the responses of those high in cognitive skills (e.g., Krosnick, 1992; Narayan & Krosnick 1996).

If there is a convention regarding the order in which response alternatives to such questions should be offered, one might presume that researchers would never violate it, so the problems we outlined above would never occur. Indeed, Guilford (1954) appears to have recognized this conversational convention, recommending that the "good" end of a rating scale be placed before the "bad" end, in order to seem most natural, particularly for unsophisticated respondents. But we have been unable to find that advice anywhere in more contemporary guidebooks for questionnaire design.

Furthermore, there is a reason researchers sometimes do not follow this recommendation: response order effects. During the last eight decades, many studies have demonstrated that the order in which response choices are offered to respondents can influence the distribution of answers to closed-ended questions, sometimes advantaging alternatives presented first, and other times advantaging alternatives presented last (for a review, see Krosnick & Fabrigar, in press). In order to minimize the impact of such response order effects on response distributions, some questionnaire design experts have advised that response order be systematically rotated across respondents (e.g., Carp, 1974; Coney, 1977; Krosnick, 1993; Matthews, 1927; Mosier & Price, 1945; Oppenheim, 1992; Rea & Parker, 1997).

For example, when two response alternatives are offered in a question, half of the respondents would be randomly assigned to receive the alternatives in one order, and the other half would receive the alternatives in the opposite order. On the assumption that respondents' answers are equally distorted in opposite directions by the two response orders, combining everyone's answers presum-

ably eliminates bias in the overall distribution of responses. At least one major survey firm, the Gallup Organization, routinely rotates response alternatives in order to estimate and control for response order effects (Moore & Newport, 1996; D. M. Moore, personal communication, April 8, 1999). The only apparent costs of such rotation have been that it increases the complexity and expense of the survey and introduces a source of systematic measurement error that must then be modeled in multivariate statistical analyses (e.g., Rea & Parker, 1997).

However, the reasoning about conversational conventions offered above suggests that such rotation may introduce as-yet unrecognized problems: distraction from thinking about the question and, consequently, reduced data quality among the respondents who receive response alternatives in unconventional orders. But if the reasoning offered above is incorrect and response order rotation does not compromise data quality, then rotation would seem to be a more appealing solution to the response order effect problem.

The research described in this paper was designed to explore these issues to gain insight into the cognitive processes of question answering. First, we assessed the conversational conventions respondents bring to bear when they are asked questions with simple, dichotomous positive and negative response alternatives. Specifically, we empirically tested the hypothesis that people expect the positive/affirmative answer choice to be presented before the negative answer choice. A second study examined whether cognitive processing is disrupted by violating this expectation, via a measure of response latency. Three other studies investigated whether violation of this convention reduces survey data quality, and another study explored the impact of violation on the contents of people's thoughts about a question.

STUDY 1

To assess the conventions governing formulation of questions offering dichotomous positive/negative response alternatives, we asked respondents to fill in blanks in hypothetical conversations, a technique previously used to identify conversational conventions (Krosnick et al., 1990).

Method

Respondents were 230 undergraduates enrolled in introductory psychology classes at the Ohio State University. Dozens of individuals completed self-administered questionnaires in large rooms. Respondents were randomly assigned to answer one of four questions.³

Each question began with the same instructions: "The following is a conversation between Bob and Pete. Please read the conversation and fill in the blanks

³ The questionnaires were distributed to students, and returning them for course credit was voluntary. The proportion of people returning each questionnaire varied, which accounts for the unequal *N*s in Table 1.

TABLE 1
Perceptions of Conversational Conventions in Study 1

Item	Answer			Total	N
	"For" in first blank/ "against" in second blank	"Against" in first blank/ "for" in second blank	Other		
Spotted owl	83.6%	13.4%	3.0%	100%	67
Mandatory insurance	93.3%	3.3%	3.3%	100%	30
Drug development	87.9%	10.1%	2.0%	100%	99
Import tax	88.2%	8.8%	2.9%	100%	34

with the word or words that you think best fit what Bob would probably have said."

One conversation involved a proposed import tax:

Import Tax

Bob: Did you know that the government is considering raising the import tax on steel?

Pete: Yeah, I read about it in the newspaper.

Bob: It's a tough issue. Automotive parts and other steel products would cost more, but there would also be more jobs for Americans. Raising the tax would help American steel industries, but it would make the cost of steel in the U.S. more expensive. What do you think? Would you vote _____ the import tax on steel, or _____ it?

Similar conversations about three other issues, limiting logging to protect the spotted owl, approving new drugs for life-threatening diseases before they are fully tested, and requiring all drivers to have accident insurance, were used. The order of the sentences in Bob's last utterance was varied across respondents. Half of the respondents, selected randomly, read Bob's last utterance as it appears above. For the other half of the respondents, the order of the second and third sentences in Bob's final utterance was reversed. So, for example, in the import tax question, these two sentences read: "Raising the tax would help American steel industries, but it would make the cost of steel in the U.S. more expensive. Automotive parts and other steel products would cost more, but there would also be more jobs for Americans."

Results

Overall, 87% of respondents filled the first blank with "for," "supporting," or synonyms of them, and filled the second blank with "against," "opposing," or synonyms of them. Only 10% of respondents filled the first blank with negative words like "against" and the second blank with positive words like "for." The remaining 3% of respondents apparently failed to understand the question, because they wrote irrelevant comments, like simply reiterating that Bob wants Pete to answer the question. As Table 1 illustrates, this pattern was apparent for each of the four items. Thus, people overwhelmingly shared the expectation that

questioners would present the favorable response alternative before the negative one in everyday conversation.

STUDY 2

Our second study tested the hypothesis that asking a dichotomous-choice question with an unconventional response-alternative order surprises respondents and temporarily distracts them. To do this, we measured how long it took respondents to answer questions with response alternatives in either the conventional or the unconventional order, expecting to see longer response latencies when people encountered the latter. To test whether the disruptive effect of conversational convention violation is greatest among respondents who are least skilled and practiced at the cognitive processes of generating answers, we divided respondents according to their cognitive skills, as measured by college grade point average (GPA). College GPA is highly correlated with other measures of cognitive skills, such as SAT scores (e.g., College Entrance Examination Board, 1997; Lawlor, Richman, & Richman, 1997), GRE scores, and the Wechsler Adult Intelligence Scale–Revised (e.g., Carvajal & Pauls, 1995).

Method

Respondents. Eighty-one Ohio State University undergraduates enrolled in introductory psychology classes visited a laboratory and answered questions on computers while seated by themselves in small rooms.

Measures. Six target items were interspersed among 26 other questions, all with simple, dichotomous alternatives (e.g., “Some people feel the federal government should take action to reduce the inflation rate, even if it means that unemployment would go up a lot. Others feel the government should take action to reduce the rate of unemployment, even if it means that inflation would go up a lot. Which do you think is more important, to reduce inflation or unemployment?”). Only the six target items were referendum questions offering “vote for” and “vote against” as the response alternatives. Some of the nontarget questions used positive and negative response alternatives (e.g., agree/disagree, favor/oppose, easier/harder), and some offered other sorts of categorical response alternatives. The import tax target item was phrased as follows:

Import Tax

The federal government is considering raising the import tax on steel that comes into the United States from other countries. Raising the steel tax would protect the steel industry from foreign competition and create more jobs for American steel workers. However, it would also increase the prices Americans pay for products made from steel. If you could vote on this, would you vote (for/against) raising the import tax on steel or would you vote (against/for) it?

Similar questions were asked about limiting logging to protect the spotted owl, approving new drugs for life-threatening diseases before they are fully tested, requiring all drivers to have accident insurance, passing stricter antipollution

laws for cars made in the United States, and restricting immigration into the United States. For each respondent, three target questions were asked in the conventional for/against order, and three were asked in the unconventional against/for order. The response-alternative order alternated between for/against and against/for from target question to target question. For half of the respondents (selected randomly), the first target question used the for/against order, and for the other half, the first target question employed the against/for order. The target questions were the 4th, 8th, 13th, 18th, 22nd, and 26th questions in the sequence. The order in which the target items were presented was randomly assigned.

Each of the target questions mentioned at least one good likely effect of the proposed policy and at least one bad likely effect of the policy. Half of the respondents (selected randomly) read the positive and negative information as shown above. For the other half of the respondents, the order in which the positive information and the negative information were presented was reversed.

Respondents reported their college GPAs up until that point in their college careers (see, e.g., Krosnick, 1992). Eight respondents failed to report their GPAs and were excluded from analyses involving cognitive skills.

Procedure. Respondents first read the following instructions explaining the procedure:

Thank you for participating in this experiment. You will be asked a series of questions about your opinions on various issues. Each question will begin with a brief introduction. After you have read the introduction, you will be asked to press the space bar. Then the question itself will be shown on the screen. To answer a question, you can push either the red key or the blue key. For each question, you will be told what answer each key corresponds to. The answer corresponding to the red key will be on the left side of the screen and will be highlighted in red. The answer corresponding to the blue key will be on the right side of the screen and will be highlighted in blue. Please keep one finger on the red key and one finger on the blue key. Push the space bar with one of your thumbs.

Because the first target item was the fourth question, respondents had several "practice" questions before being asked to answer a target question. Therefore, they were likely to have mastered the procedure by the first target item.

For each question, background information first appeared on the screen. The background information for the import tax question consisted of the first three sentences of this question. These appeared on the computer screen first with instructions at the bottom of the screen to press the space bar when finished. When respondents pressed the space bar, half of them (selected randomly) saw the question with the response alternatives in the for/against order. The other half of the respondents saw the question with the response alternatives in the against/for order. In the bottom left and right corners of this screen were the choices "for" and "against" in the order presented in the question. The choice at the bottom left was in red, and the choice at the bottom right was in blue. The "1" key on the computer keyboard was colored red, and the "9" key was colored blue, corresponding to the colors on the screen.

The computer recorded the amount of time that passed between the moment a respondent pressed the space bar to see the question and the moment when he or she pressed the red or blue key. If a respondent pushed some other key, his or her response time for that question was recorded as invalid. Only three such errors were recorded in the entire experiment.

Response time preparation. Following Fazio's (1990) recommendation, the response times were subjected to a reciprocal transformation.

Results

As expected, the average response time for questions asked with the conventional response order was shorter (3.67 s) than that for questions asked with unconventional response order (4.04 s). According to a repeated-measures analysis of variance, the effect of response order on response time was statistically significant ($F(1, 80) = 3.83, p < .05$).⁴ And as expected, this effect was much larger and highly significant among respondents with GPAs of 3.0 or less: .79 s ($F(1, 35) = 4.36, p = .01$). In contrast, the effect was much smaller, .12 s, and not significant among respondents with GPAs higher than 3.0 ($F(1, 36) = 0.47, ns$). The interaction between college GPA and response order was marginally significant ($F(1, 71) = 2.08, p < .10$).

Only one of the six items, about mandatory insurance, manifested a significant effect of response choice order on the distribution of people's answers ($\chi^2(1) = 5.74, p < .01$). When the "for" alternative appeared first, 32% of respondents selected it, compared to 58% who selected "for" when it appeared second. This constitutes a recency effect of 26%.

The interaction between response order and GPA was significant ($t(73) = 1.85, p < .05$). In line with the notion that response order effects are concentrated among people low in cognitive skills (Krosnick, 1991), this effect was much larger (17% selected "for" when it was presented first and 67% when it was presented last, a recency effect of 50%) and was highly significant among respondents with GPAs of 3.0 or less ($\chi^2(1) = 9.26, p < .001$), whereas it was much smaller (44% selected "for" when it was presented first and 52% when it was presented last, a recency effect of 8%) and was nonsignificant among respondents with GPAs higher than 3.0 ($\chi^2(1) = 0.27, ns$). No other items manifested significant response order effects among respondents in either the low- or the high-GPA groups when analyzed separately.

STUDY 3

The results of Study 2 are consistent with the claim that respondents were surprised and distracted by the unconventional wordings. However, these data are also consistent with the claim that the convention violation did not distract people at all, but rather made them more focused and thoughtful in answering the

⁴ This hypothesis and most of those tested later are clearly directional, so one-tailed significance tests are reported unless otherwise indicated.

questions. That is, the longer response times for unconventionally worded questions could indicate more thoughtful consideration in generating responses.

If in fact unconventional ordering disrupted respondents, we would expect the quality of data to go down. That is, if surprise caused by convention violation temporarily distracted respondents from the task of question answering, they may have returned to that task a bit less likely to thoroughly canvass their memories for relevant information and a bit less likely to carefully integrate what they retrieved into summary judgments. As a result, the reports they provided may have been slightly less accurate reflections of their dispositions on the issues in question, especially among respondents lowest in cognitive skills. But if the convention violation instead increased focused thought about the question topic, response quality might have increased, leading responses to have been more accurate reflections of people's dispositions. Our next study tested these competing predictions in a general population survey, which employed a proxy measure of cognitive skills that is very strongly correlated with more direct measures: amount of formal education (see Ceci, 1991).

Specifically, we compared the ability of various attitudinal and demographic variables to predict responses to questions asked with the conventional response order and responses to the same questions asked with the unconventional response order. If responses to the conventionally asked questions were more highly correlated with the predictors than responses to questions asked with the unconventional order, this would support the hypothesis that responses to the unconventionally asked questions were the product of less thoughtful consideration.

Method

Respondents. For this study, the Roper Organization included four questions in an omnibus survey conducted during August 1993. A multistage, stratified probability sample of 1982 American adults was interviewed face-to-face.⁵

Measures of target attitudes. Four of the target items used in Study 2 (on the import tax, the spotted owl, mandatory insurance, and drug development) were interspersed among approximately 150 other questions asked in this survey. The import tax item was the 28th question in the interview; the spotted owl item was the 35th question; the drug development item was the 97th question; and the mandatory insurance item was the 111th question.

For half of the respondents, the spotted owl and mandatory insurance items were presented with the "for" alternative first, and the import tax and drug

⁵ Quotas were specified based upon gender, age, and employment status using U.S. Census population data. Forty-seven percent of the sample respondents were male, and 53% were female. Eighty-two percent were White, 13% were Black, 3% were Asian, and 2% were other races. Twenty percent of the sample had less than a high school education, 37% were high school graduates, and 43% had at least attended some college. Fifty-six percent of respondents were married, and 48% were between ages 18 and 39. The Roper Organization cannot report a response rate for the survey, because records of refusals are not kept for quota surveys of this sort.

development items offered the "against" alternative first. For the other half of the respondents, the import tax and drug development items presented the "for" alternative first, and the spotted owl and mandatory insurance items presented the "against" alternative first. Each interviewer alternated administering the two versions of the questionnaire across his or her sequence of respondents. To be sure that the random assignment procedure used was effective, we compared the two half-samples of respondents in terms of various demographics and attitudes, including education, age, gender, union membership, income, political party identification, and political ideology. No statistically reliable differences appeared, suggesting that the random assignment had been adequately implemented.

Attitudinal dispositions. In order to gauge the degree to which respondents' referendum answers reflected their dispositions, we used other questions in the survey measuring such dispositions. Respondents were asked: "Change can be started from many difference sources. Reading down the list in front of you, which groups do you feel have caused the most significant changes *for the better* in our society in the past few decades? Which groups have caused the most significant changes *for the worse*?" Among this list of groups were environmentalists, scientists and inventors, and government leaders. Respondents who said a group had caused changes for the worse were given a value of 0 for that group; respondents who did not mention a group as causing changes for the better or worse were given a value of .5; respondents who said a group had caused changes for the better were given a score of 1.

We thought that opinions on the spotted owl might be based upon attitudes toward environmentalists, who advocate protection of endangered species. We thought that opinions on the drug development issue might be based on attitudes toward scientists and inventors, because they are presumably responsible for most drug development. And we suspected that responses to all the target items might be based partly on attitudes toward government leaders, because all the proposed policies would be implemented by the federal government. We also thought that all opinions might be predicted by political ideology (coded 0, 0.25, 0.50, 0.75, and 1.0 for people who were very conservative, moderately conservative, middle of the road, moderately liberal, and very liberal, respectively) and party identification (coded 0 for Republicans, .5 for Independents, and 1.0 for Democrats).

Demographics. Demographic characteristics for each respondent were also measured, including race, education, gender, and age. Race was coded 1 if the respondent was White and 0 otherwise. Education was coded 0 if the respondent had no formal schooling, 0.17 if he or she attended only grade school, 0.33 if he or she attended some high school but did not get a degree, 0.5 if he or she was a high school graduate, 0.67 if he or she had some college education but did not get a degree, 0.83 if he or she was a college graduate, and 1 if he or she had completed at least some postgraduate work. This coding scheme assigns equally spaced values to the seven categories used by Roper to describe respondent

education. Gender was coded 1 for males and 0 for females. Age was coded 0 for people ages 18 to 20, .1 for people 21–24, .2 for people 25–29, .3 for people 30–34, .4 for people 35–39, .5 for people 40–44, .6 for people 45–49, .7 for people 50–54, .8 for people 55–59, .9 for people 60–64, and 1 for people 65 and older. Again, this coding scheme assigns equally spaced values to the categories used by Roper to describe respondent age.

Results

Data quality. To assess the degree to which people's attitude reports reflected their dispositions, we estimated the parameters of regression equations predicting each attitude report (coded 1 if the respondent voted for the policy and 0 if he or she voted against it) using the attitudinal dispositions and demographics. The parameters of each regression equation were estimated separately for respondents who received the for/against response order and those who received the against/for order.

As Table 2 illustrates, significant predictors of each target item appeared, and these predictors varied across topics and response orders.⁶ When the for/against response order was used, support for the spotted owl issue was highest among liberals, people with positive attitudes toward environmentalists, people with more education, and females. Liberals, people with a positive attitude toward government leaders, Whites, older people, and highly educated respondents were more likely to support a mandatory insurance law. Support for expediting drug development was especially prominent among liberals, people who were positive toward scientists/inventors, and highly educated respondents. Whites were especially supportive of import taxes. Thus, 13 significant or marginally significant effects appeared.⁷

Every one of these relations became weaker when respondents were asked the against/for question instead. Specifically, they dropped from an average absolute value of .13 to an average absolute value of .06, and only 3 of the 13 predictors manifested significant relations with responses to the against/for questions. In four cases, a variable was a significant predictor of responses to the against/for item that had not been a significant predictor of answers to the for/against item. In total, only 7 significant or marginally significant effects appeared when the against/for response order was used, as compared to the 13 that appeared when the for/against order was used. And the average absolute value of the 30 coefficients predicting responses to the for/against items was greater (0.07) than the average absolute value of the 30 coefficients predicting answers to the against/for items (0.05).

To more formally test the notion that answers to the latter items were less predictable than answers to the former items, we compared the R^2 s for equations

⁶ Because our hypotheses about the significance of these coefficients were nondirectional, two-tailed tests of significance were used here.

⁷ The variance of each predictor was virtually identical across the groups.

TABLE 2
Unstandardized Regression Coefficients from Equations Predicting
Referendum Question Answers in Study 3

Predictors	Target item							
	Spotted owl		Mandatory insurance		Drug development		Import tax	
	For/ against order	Against/ for order	For/ against order	Against/ for order	For/ against order	Against/ for order	For/ against order	Against/ for order
Attitude toward environmentalists	.19**	.14*	—	—	—	—	—	—
Attitude toward scientists, inventors	—	—	—	—	.09†	-.01	—	—
Ideology	.23**	.12	.13†	-.04	.25**	.07	.05	.04
Party identification	.00	.11*	-.02	.02	-.07	.00	.02	.06
Attitude toward government leaders	.02	-.01	.16**	.15**	.02	-.09†	.02	-.01
White	.00	-.02	.10*	.06	-.05	.02	.20**	.06
Education	.04**	.04*	.04*	.02	.03*	.02	.01	-.03*
Gender	-.06†	.00	.05	.00	.01	.08*	.01	-.00
Age	-.09	-.01	.14*	.01	-.02	.09	.03	-.05
R^2	.06	.03	.04	.02	.03	.02	.02	.01
N	759	743	768	745	777	782	718	727

† $p < .10$.

* $p < .05$.

** $p < .01$.

for the for/against questions to those for the against/for questions (see Table 2). The penultimate row of Table 2 displays the R^2 s, which were larger for the for/against questions than the against/for questions for all four issues. Collapsing across the four items (see Rosenthal & Rubin, 1986), the average R^2 for the for/against questions was significantly larger than the average R^2 for the against/for questions ($t(792) = 1.91, p < .05$; average effect size $g = .10$). This difference was larger among respondents with a high school education or less ($g = .13$) than among respondents with more than a high school education ($g = .09$), and the difference was statistically significant in the former group ($t(392) = 1.69, p < .05$), but not in the latter ($t(392) = 1.18, ns$), although the interaction between response order and education was not significant (focused comparison of effect sizes: $z = .36, ns$).

Effects on distributions of "for" and "against" answers. As in Study 2, only one of the four target items manifested a significant effect of response choice order on the distribution of answers. For the import tax item, 60.6% of respondents said they would vote in favor when that alternative was presented first, and 65.0% said they would do so when that alternative was presented second. Thus, a recency effect of 4.4% occurred. When entered as a predictor of respondents'

answers with the attitudinal dispositions and demographic characteristics shown in Table 2, response order was a significant predictor of answers to the import tax item ($t(1444) = 1.97, p < .05$).

Also consistent with Study 2, the interaction of question form with education was statistically significant ($t(1443) = 2.01, p < .05$). The recency effect was much larger and statistically significant among respondents with a high school education or less ($t(755) = 2.38, p < .01$). Among these respondents, 61.9% said they would vote in favor when that alternative was presented first, and 68.8% said they would do so when that alternative was presented second, a recency effect of 6.9%. In contrast, the effect was much smaller and was nonsignificant for respondents with more than a high school education ($t(688) = 0.18, ns$). Among these respondents, 58.9% said they would vote in favor when that alternative was presented first, and 60.7% said they would do so when that alternative was presented second, a recency effect of 1.8%. No response order effects appeared for any of the other three attitude questions when the education subgroups were each examined individually.

STUDY 4

Study 3's results are consistent with the hypothesis that the surprise and distraction caused by convention violation led responses to less accurately reflect respondents' dispositions on the issues in question. This was most apparent among people lowest in cognitive skills, for whom recovery from distraction was presumably most challenging. However, we cannot be sure that the dispositions measured in Study 3 did in fact reflect optimal attitude reporting. In principle, it is possible that some or all of the relations observed were evidence of superficial decision making rather than thoughtful and valid attitude reporting.

Our next study was designed to yield more unambiguous evidence about the quality of attitudinal reports. To do so, we made use of a research method developed by economists called "contingent valuation" (Carson & Mitchell, 1995). The purpose of this sort of research is to gauge the economic value that people attach to "public goods," such as parks and undeveloped sites of natural beauty. Some such studies tell respondents about an event that caused a natural resource to be damaged, and respondents are asked to indicate how much money they would be willing to pay to prevent further such damage from occurring. This amount is taken to be a lower bound estimate of the value of the existence of the resource.

A central principle in this research, and in fact in all of economics, is that people should not place a lower value on a greater quantity of a particular good (Carson & Mitchell, 1995). So, for example, if people are asked to value the existence of a 1-mile segment of clean river water or a 10-mile segment, they should place no lower value on the latter than on the former. It is certainly possible that a 1-mile segment is sufficient to serve people's needs and desires, so they may place no greater value on a 10-mile segment. But as a general rule,

people are expected to attach more value to a greater quantity of a particular public good.

If employing an unconventional ordering of response choices in a question measuring the value of a public good surprises and distracts respondents and lowers the quality of their responses, then these responses may be less reflective of the size of the good being valued. To test this prediction, our next study described a human-caused incident that damaged a natural resource. Some respondents were told that the damage was of greater magnitude than other respondents. We expected this manipulation to have less impact on the responses of people given an unconventional response choice order than on the answers of people given a conventional response choice order, especially among people with more limited cognitive skills.

Method

Respondents. The 794 respondents in this study participated in telephone interviews conducted by the Ohio State University Survey Research Unit between November 4 and December 23, 1998. A sample of randomly generated telephone numbers representing all households in the state of Ohio were called, and within a household, an English-speaking adult was randomly chosen to be interviewed.

Measures. Toward the end of a 20-min interview, respondents were told at length about PCBs deposited in an Ohio river or multiple such rivers, the effects these PCBs were having on wildlife, and a proposed plan to cover up the PCBs and prevent them from having effects in the future. Respondents were told that if the plan were to be adopted by the state, each household in Ohio would pay a special one-time tax to cover the costs of implementing the plan. Finally, respondents were asked to indicate whether they would vote for or against the proposed plan if it were offered as a referendum on a ballot at a specified cost per Ohio household. Half of respondents were randomly assigned to be asked whether they would vote "for or against" the plan, and the other half were asked instead whether they would vote "against or for" the plan.

Respondents were independently randomly assigned to be told that the PCBs were deposited either in one river totaling 39 miles long or in four rivers totaling 317 miles long. More people should presumably have been willing to support the plan if it would cover PCBs in four rivers than if it would cover PCBs in only one river for the same cost. We explored whether answers to the against/for question would be less responsive to this manipulation than answers to the for/against question.

Educational attainment again served as a proxy measure of cognitive skills.

Results

Data quality. Among all people asked the for/against question, the manipulation of the number of rivers had a significant effect on responses ($t(357) = 2.05$, $p < .05$). Eighty percent of people said they would vote for the plan when it

would cover PCBs in four rivers, while only 70% said they would vote for the plan when it would cover PCBs in only one river. But among all people asked the against/for question, the manipulation had no effect. Sixty-six percent of respondent said they would vote for the plan when it would cover PCBs in four rivers, and 65% said they would vote for the plan when PCBs would be covered in only one river, a nonsignificant difference ($t(354) = 0.30, ns$). However, the impact of the number of rivers did not differ significantly between the two response choice orders for all respondents combined ($z = 1.14, ns$).

Among respondents with a high school education or less, responsiveness to the manipulation of number of rivers differed even more substantially between people asked the for/against question and people asked the against/for question ($z = 1.80, p < .05$). Among people asked the for/against question, 85% said they would vote for the plan when it would cover PCBs in four rivers, and 69% said they would vote for the plan when it would cover PCBs in only one river ($t(153) = 2.36, p = .01$). In contrast, among people who received the against/for question, 69% said they would vote for the plan when it would cover PCBs in four rivers, and 71% said they would vote for the plan when it would cover PCBs in only one river, a nonsignificant difference ($t(145) = 0.30, ns$). Thus, among the least educated respondents, responsiveness to the number of rivers was greater under the for/against order than under the against/for order.

In contrast, among respondents with more than a high school education, the effect of response choice order on responsiveness to the manipulation did not approach significance ($z = 0.11, ns$). When these respondents were asked the for/against question, 76% said they would vote for the policy when it would cover PCBs in four rivers, and 71% said they would vote for it when it covered PCBs in one river, a difference of 5% ($t(201) = 0.80, ns$). When asked the against/for question, 65% of these respondents said they would vote for the policy when it covered PCBs in four rivers, and 60% said they would vote for it when it covered PCBs in one river, again a difference of 5% ($t(206) = 0.68, ns$). Thus, among highly educated respondents, the impact of the manipulation of number of rivers was unaffected by response order. However, the impact of response order on responsiveness to the manipulation did not differ significantly between the two education groups ($z = 1.25, ns$).

Effects on distributions of "for" and "against" answers. Response choice order significantly affected the distribution of answers. Seventy-five percent of respondents said they would vote for the plan when that alternative was presented first, and 66% said they would do so when that alternative was presented second. Thus, a significant primacy effect of 9% occurred ($t(711) = 2.72, p < .01$). The interaction of response choice order with education was not statistically significant ($t(710) = .86, ns$).

STUDY 5

Taken together, Studies 3 and 4 are consistent with the notion that cognitive disruption caused by conversational convention violation reduces the quality of

attitudinal reports. We have presumed that this disruption was due to respondent surprise and distraction. However, there is another possible explanation as well. People usually approach the task of answering questions with a confirmatory bias (Hoch, 1984; Koriat, Lichtenstein, & Fischhoff, 1980; Wason & Johnson-Laird, 1972; see also Snyder & Gangestad, 1981). So for the sorts of questions we asked, people might generally be inclined to think of reasons to vote for a proposed policy, rather than thinking of reasons to vote against it. This tendency might even be encouraged by placing the affirmative response alternative first.

Presenting the negative alternative first might instead call respondents' attention to it and cause them to begin thinking by generating reasons to vote against the proposal. The longer reaction time apparent for responses to such questions may reflect people's relative unfamiliarity with executing disconfirmatory reasoning. Or the longer reaction time may occur because people execute their customary confirmatory reasoning after some disconfirmatory reasoning, yielding more total time spent considering possible responses. If either of these two possibilities is true, respondents would be generating more reasons to vote against a policy when the negative alternative was presented first.

Our fifth study therefore sought additional evidence of the cognitive mechanisms of the effects observed thus far. In this study, respondents were asked to list the thoughts they had when thinking about the referendum-style questions used in our previous studies. If employing the against/for order of response choices causes surprise and distraction, it should presumably lead to less question-focused thinking and more tangential thinking. This thinking could entail explicit attributional speculation about the reasons that the questioner violated the conversational convention, or this thinking could be completely irrelevant to the question, reflecting the fact that less skilled respondents' cognitive focus on the question itself was disrupted, temporarily unsettling them enough for their thoughts to wander elsewhere for just a moment. If, on the other hand, the against/for response choice order inspires more disconfirmatory thinking, it should lead to more thoughts about disadvantages of the proposed policy and perhaps fewer thoughts about its advantages.

Method

Respondents. For this study, 210 Ohio State University undergraduates enrolled in introductory psychology classes completed self-administered paper-and-pencil questionnaires in large rooms in groups of up to 17 people.

Measures. The first page of the questionnaire provided the following instructions to respondents:

In this experiment, you will read several questions about your opinions on various issues. After reading each question, please DO NOT answer it. We want to know all the thoughts that go through your mind as you think about the question and formulate your answer.

Respondents then turned the page to read one of the four target items used in Study 3. Each respondent was randomly assigned to one of the target items, and

TABLE 3
Mean Number of Times Each Type of Thought Was Listed by Respondents in Study 5

Type of thought	Response choice order		Difference
	For/against	Against/for	
Desirable effects of the policy	.20	.31	.11
Undesirable effects of the policy	.34	.33	-.01
Favorable evaluations of the policy	.67	.64	-.03
Unfavorable evaluations of the policy	.47	.47	.00
Expressions of uncertainty about the policy	.14	.11	-.03
Questions about the effects of the policy	.67	.52	-.15
Other thoughts about the policy	.94	.86	-.08
Thoughts about the question	.04	.02	-.02
Irrelevant thoughts	.28	.37	.09*
Total number of thoughts	3.75	3.57	-.18
Number of cases	100	110	

* $p < .05$.

he or she was also randomly assigned to receive the response alternatives either in the for/against order or in the against/for order. Beneath the question was the instruction: "List all your thoughts."

Respondents were also asked to report their college grade point averages.

Coding. Coding of the thoughts was done in two phases by a person who did not know anything about the hypotheses being tested. During the first phase, the coder "chunked" each respondent's answer into discrete thoughts, defined as "ideas that can stand alone." During the second phase, the coder assigned each thought to one of nine categories:

1. Desirable effects of the policy.
2. Undesirable effects of the policy.
3. Favorable evaluations of the policy.
4. Unfavorable evaluations of the policy.
5. Expressions of uncertainty about the policy.
6. Questions about the effects of the policy.
7. All other thoughts about the policy.
8. Thoughts about the question respondents were asked.
9. Irrelevant thoughts.

To gauge the reliability of the coding, a second coder chunked and coded the thoughts listed by a portion of the respondents. The coders agreed perfectly on 80% of the thought frequencies.

Results

The average number of thoughts in each of the categories offered by respondents is presented in Table 3, separately for respondents whose answer choices

were in the for/against order and for those whose answer choices were in the against/for order. The last column of Table 3 displays the differences between these two groups. Because the numbers of thoughts described in Table 3 were extremely skewed, rather than being normally distributed, we conducted statistical tests of the differences between the two experimental groups via Poisson regression (King, 1988). In these regressions, we controlled for dummy variables representing the four different questions that respondents were asked, as well as respondents' college GPAs, which were measured as in Study 2.

The frequency of most types of thoughts was unaffected by response choice order. Most importantly, there were no significant differences in the frequency of favorable or unfavorable evaluations of the policy or of thoughts about desirable or undesirable effects of the policy. This challenges the claim that the against/for response choice order inspires more disconfirmatory thinking. Indeed, although not a significant difference, there were more *confirmatory* thoughts among people presented with the against/for order than among the people presented with the for/against order. This trend also contradicts the claim that the against/for response choice order enhances disconfirmatory reasoning.

Only one category of thoughts was significantly influenced by response choice order: irrelevant thoughts. People given the against/for order offered .37 such thoughts on average, compared to .28 on average for people given the for/against order, a significant difference of .09 ($z = 2.05, p < .05$).⁸ Furthermore, this difference was much more pronounced among respondents with GPAs less than 2.8 (difference = .28, $z = 2.36, p < .05, N = 103$) than among respondents with GPAs of 2.8 or higher (difference = .03, $z = 0.54, ns, N = 107$), although the interaction between response choice order and GPA was not significant ($z = 0.29, ns$). These findings are therefore consistent with the notion that the against/for response order distracted people from thinking about the topic of the question.

Respondents very rarely voiced thoughts about the phrasing of the question presented on the questionnaire: The average frequency of this type of thought was .04 among people given the for/against order and .02 among people given the against/for order ($z = 0.98, ns$). And none of these comments explicitly mentioned the order of the response choices. Thus, we see no direct evidence here that the against/for response choice order led to more thinking about the question itself or to attributional reasoning about the convention violation. This suggests that the conversational convention violation had its effects completely outside of consciousness or so fleetingly that respondents were not able to recall thoughts about the violation even just moments later when listing those thoughts.

STUDY 6

All of the previous studies focused on referendum-style questions asking respondents whether they would "vote for" or "vote against" a proposed policy.

⁸ This significance test is two-tailed.

Our final study employed a design similar to that used in Study 3 to see whether our findings generalize to other sets of simple, dichotomous, mutually exclusive affirmative and negative response alternatives.

Method

Respondents. For this study, we analyzed data from two telephone surveys conducted by the Gallup Organization, one with a nationally representative sample of 1009 American adults conducted during March, 1998, and the other involving interviews with a national sample of 1007 adults conducted during April, 1998.

Measures of target attitudes. Two target attitudes were measured in the March, 1998, survey:

Internet

Overall, do you think that the Internet has had more of a (positive/negative) impact on society or more of a (negative/positive) impact on society?

Microsoft

Overall, do you think that Microsoft has had more of a (positive/negative) impact on the computer industry or more of a (negative/positive) impact on the computer industry?

The Internet item was the 13th question in the interview, and the Microsoft item was the 17th question in the interview. For half of the respondents (selected randomly), these items were presented with the “positive” response alternative first, and for the other half of respondents, the “negative” response alternative was presented first.

Two other target attitudes were measured in the April, 1998, survey:

Company Mergers

Still thinking about corporate mergers, when two large companies merge and create one VERY large company, do you think the result is usually (*good/bad*) for consumers or (*bad/good*) for consumers?

Bank Mergers

Still thinking about corporate mergers, when two large banks merge and create one VERY large bank, do you think the result is usually (*good/bad*) for consumers or (*bad/good*) for consumers?

Half of respondents (selected randomly) were asked the company mergers question, and the other half were asked the bank mergers question. For half of respondents (selected randomly) who answered the company mergers question, the “good” response alternative was presented first, and the “bad” response alternative was presented second. For the other half of these respondents, the “bad” response alternative was presented first. Similarly, for half of respondents (selected randomly) who answered the bank mergers question, the “good” response alternative was presented first, and the “bad” response alternative was

presented second. For the other half of these respondents, the “bad” response alternative was presented first.

Attitudinal and behavioral dispositions. We again gauged the degree to which respondents’ answers reflected their dispositions. Specifically, we thought that Internet and computer use and attitudes toward Microsoft and Bill Gates (founder and then-CEO of Microsoft) would be relevant to responses to both the Internet and the Microsoft items in the March survey, and that using Microsoft Internet Explorer, beliefs about whether Microsoft was a monopoly, and following news about investigations of Microsoft would be relevant to the Microsoft item.

Respondents in the March survey were asked whether they used a computer at home and at work and whether they used the Internet or other online services from home or work. Each of these four items was coded 1 if a respondent said “yes” and 0 if he or she said “no.” The four items were then averaged into an index of computer and Internet use.

Respondents in this survey were also asked whether they felt favorable or unfavorable toward Microsoft and Bill Gates. Answers to each of these items were coded 1 if a respondent felt favorable, 0 if he or she felt unfavorable, and .5 if he or she felt neutral or mixed.

Respondents in this survey were asked whether they thought Microsoft was a monopoly, and whether or not they used Microsoft Internet Explorer (the Web browser produced by Microsoft). Responses to each of these items were coded 1 if a respondent said “yes” and 0 if he or she said “no.”

Finally, respondents in this survey were asked how closely they had been following the news about Justice Department and Congressional investigations into Microsoft’s business practices. Responses were coded 0 if a respondent said “not at all,” .33 if a respondent said “not too closely,” .67 if a respondent said “somewhat closely,” and 1 if a respondent said “very closely.”

With regard to the mergers questions in the April survey, we thought other opinions about mergers and companies might be relevant. Respondents in that survey were asked whether the government should allow mergers between two large companies to happen or whether the government should do more to stop these mergers. Responses were coded 1 if a respondent said the government should allow such mergers and 0 if he or she said the government should do more to stop them. Respondents were also asked whether they agreed that there is too much power concentrated in the hands of a few big companies. Responses were coded 0 if a respondent completely agreed, .33 if he or she mostly agreed, .67 if he or she mostly disagreed, and 1 if he or she completely disagreed.

We also thought that all four target attitudes might be predicted by political ideology (coded 0, 0.25, 0.50, 0.75, and 1.0 for people who were very conservative, moderately conservative, middle of the road, moderately liberal, and very liberal, respectively) and party identification (coded 0 for Republicans, .5 for Independents, and 1.0 for Democrats).

Demographics. Demographic characteristics of each respondent were also measured, including race, education, gender, and age. Race was coded 1 if the

respondent was White and 0 otherwise. Education was coded 0 if the respondent had between a 1st- and a 4th-grade education, 0.13 if he or she had between a 5th- and a 7th-grade education, 0.25 if he or she had an 8th-grade education, 0.38 if he or she had between a 9th- and an 11th-grade education, 0.50 if he or she had graduated from high school, 0.63 if he or she had some technical education beyond high school, .75 if he or she had some college education, .88 if he or she was a college graduate, and 1 if he or she had completed at least some postgraduate work. Gender was coded 1 for males and 0 for females. Age in years was recoded so that the minimum age (18) was coded as 0 and the maximum age (99) was coded as 1.

Results

Data quality. To assess the degree to which people's attitude reports reflected their dispositions, we estimated the parameters of regression equations predicting each target attitude report (coded 1 if the respondent answered "good" or "positive" and 0 if he or she answered "bad" or "negative") using the demographics and relevant attitudinal and behavioral dispositions for each item. The parameters of each regression equation were estimated separately for respondents who received the positive/negative response order and those who received the negative/positive order.⁹

Collapsing across the four target attitudes, the average R^2 for the positive/negative items (0.31) was significantly larger than the average R^2 for the negative/positive items (0.20; $t(481) = 1.70, p < .05$; average effect size $g = 0.17$).¹⁰ This difference was large and statistically significant among respondents with a high school education or less ($g = .35; t(71) = 1.98, p < .01$; see Table 4), and small and nonsignificant among respondents with more than a high school education ($g = .05; t(101) = 0.31, ns$). The interaction between response choice order and education was marginally significant (focused comparison of effect sizes: $z = 1.36, p < .10$).

Effects on distributions of answers. None of the four target items manifested a significant effect of response choice order on the distribution of answers.

META-ANALYSIS

Although the results of these studies are all consistent with the notion that the effect of response order on cognitive processing is strongest among respondents with the least cognitive skills (e.g., those with the least education in Studies 3, 4, and 6 and with low college GPAs in Studies 2 and 5), the differences between respondents high and low in cognitive skills were marginally significant in

⁹ The variance of each predictor was virtually identical across the groups.

¹⁰ This statistic is based on a comparison of the t -statistics for the positive/negative regression equations with those for the negative/positive regression equations. Because these t -statistics are a function of the R^2 value for the equation and the degrees of freedom, this statistic controls for unequal sample sizes.

TABLE 4
Unstandardized Regression Coefficients from Equations Predicting Answers
from Low-Education Respondents in Study 6

Predictors	Target item							
	Internet		Microsoft		Mergers between large companies		Bank mergers	
	Positive/ negative order	Negative/ positive order	Positive/ negative order	Negative/ positive order	Positive/ negative order	Negative/ positive order	Positive/ negative order	Negative/ positive order
Computer and Internet use	.35**	.24	.17	-.04	—	—	—	—
Computers are an important part of your life	-.07	.14	-.04	.05	—	—	—	—
Favorable toward Microsoft	.34*	.08	-.04	.01	—	—	—	—
Favorable toward Bill Gates	-.12	-.12	.28**	.21**	—	—	—	—
Microsoft has a monopoly	—	—	-.16*	.00	—	—	—	—
Use Microsoft Web browser	—	—	-.10	-.02	—	—	—	—
Attention to Microsoft investigation	—	—	-.13	-.04	—	—	—	—
Too much power is in the hands of a few businesses	—	—	—	—	.46**	.30	-.21	.03
The government should allow mergers between large companies	—	—	—	—	.54**	.39**	.42**	.35**
Ideology	.04	-.07	.13	-.05	.28	.12	.18	.03
Party identification	-.08	.16	.01	-.05	-.09	-.08	-.24†	-.27*
White	.13	.11	.13	.02	-.03	.04	-.11	-.02
Gender	.13	.14†	.10	-.04	-.06	-.08	-.15	.00
Age	-.41†	-.18	-.27	-.30*	.01	-.07	-.47*	-.33†
R^2	.20	.12	.24	.21	.45	.22	.36	.26
N	141	134	123	121	90	89	76	98

† $p < .10$.

* $p < .05$.

** $p < .01$.

Studies 2, 4, and 6 and nonsignificant in Studies 3 and 5. To test whether this effect is in fact reliable, we meta-analyzed the effect of response order for respondents low and high in cognitive skills using the statistics shown in Table 5.

For Studies 2, 3, 5, and 6, these statistics represent the effect of response order on response latency, R^2 , or listing of irrelevant thoughts. For Study 4, these statistics represent the interaction between response order and the effect of the number of rivers in which PCBs would be covered. As shown in Table 5, the average Cohen's d effect size was .41 for respondents with more limited cognitive skills (a "moderate" effect size; Cohen, 1988), whereas it was only .10

TABLE 5
Effects of Response Order Included in the Meta-analysis

	Test statistic	<i>N</i>	Cohen's <i>D</i>
<u>Study 2</u>			
Low cognitive skills	$F(1, 35) = 4.36$	36	.71
High cognitive skills	$F(1, 36) = 0.47$	37	.23
<u>Study 3</u>			
Low cognitive skills	$t(392) = 1.69$	400	.17
High cognitive skills	$t(392) = 1.18$	400	.12
<u>Study 4</u>			
Low cognitive skills	$t(298) = 1.79$	302	.21
High cognitive skills	$t(407) = 0.05$	411	.01
<u>Study 5</u>			
Low cognitive skills	$z = 2.36$	103	.48
High cognitive skills	$z = 0.54$	107	.10
<u>Study 6</u>			
Low cognitive skills	$t(71) = 1.98$	81	.47
High cognitive skills	$t(101) = 0.31$	111	.06
<u>Averaged across studies</u>			
Low cognitive skills			.41
High cognitive skills			.10

(a "small" effect size; Cohen, 1988) for respondents with greater cognitive skills. A meta-analysis combining across the studies showed that the interaction between cognitive skills and cognitive disruption was significant (focused comparison of effect sizes: $z = 2.22$, $p = .01$).

DISCUSSION

Conversational Conventions and Response Processing

These results support the hypothesis that people believe it is conventional to ask questions with simple, dichotomous positive and negative response alternatives with the affirmative alternative first. When this convention was violated, it disrupted respondents' cognitive processing (Studies 3 and 4), slowing down their responses (Study 2) and yielding less predictable reports, presumably containing more error (Studies 3, 4, and 6). These effects were significant only among respondents with the least skill in executing cognitive processes diligently

(indicated here by low GPAs and low levels of educational attainment). Thus, cognitive disruption was greatest among individuals whose abilities to stay focused on the task were the most limited.

Our findings in this regard may have implications for reinterpreting other, previously documented effects of conversational norms violation (e.g., Dulany & Hilton, 1991). Consider, for example, the dilution effect (Nisbett et al., 1981). Tetlock et al. (1996) showed that presenting nondiagnostic information causes underuse of diagnostic information partly because people assume that speakers are conforming to a norm dictating that all provided information should be relevant to the issue at hand. Consequently, people presumably search for a way that the nondiagnostic information might be relevant and then incorporate it in their judgment, reducing the impact of other considerations.

Our findings suggest another possibility: that some people may immediately recognize nondiagnostic information as such and may label it as a puzzling violation of a conversational norm. They may therefore think not about how the information might be relevant to the judgment task but rather about why the researchers would have so obviously violated a conversational norm. The resulting confusion may distract some respondents from the target question momentarily, and they may later answer that question slightly less precisely as a result, thus reducing the correlation between diagnostic predictive information and the judgment rendered. In light of this logic, it may be useful for future research to more directly measure the mechanisms underlying apparent effects of conversational norm violations to see if distraction from the task at hand plays a role in reducing data quality.

Response Order Effects and Implications

Previous portrayals of the cognitive processes underlying response order effects anticipates few such effects for questions involving simple, dichotomous, mutually exclusive positive and negative response alternatives (e.g., Krosnick, 1991). This is because the cognitive burden of thinking about such alternatives is much less than that posed when a long list of alternatives is presented (e.g., "Which is the most important problem facing the country today: crime, drug abuse, pollution, unemployment, or inflation?"). In line with this reasoning, the only previously published tests of such response order effects found no impact on answers to agree/disagree (Schuman & Presser, 1981, p. 210, footnote 4) and yes/no questions (Jacobs, 1972). Likewise, no response order effects were found for 12 of the 15 items examined here.

However, we did find significant response order effects for 3 of the 15 items, suggesting that response order effects are possible in questions like these. These significant effects constitute the first evidence that researchers should be aware of the possibility that response order can affect answers even to questions involving simple, dichotomous, mutually exclusive positive and negative response alternatives. As we explained earlier, response order effects like these have led some researchers to believe that response choice order should be rotated across

respondents to eliminate its impact on observed distribution of responses. However, the evidence reported here suggests that doing so makes some respondents' cognitive tasks more difficult and may compromise data quality. Consequently, the best solution may be to use only the conventional response order and take steps to eliminate response order effects by enhancing respondent motivation to provide high-quality answers (see Krosnick, 1991).

More broadly, our research suggests that similar consequences could follow if other conversational conventions regarding response alternative order are violated. Although the present research focused on positive and negative response alternatives, Cooper and Ross (1975) identified a broad range of word order conventions, as we reviewed earlier. Our findings suggest that when writing questions using sets of words governed by these conventions, researchers should conform to them, because violating them may reduce data quality. We look forward to future research testing this possibility.

CONCLUSION

The notion of conversational norms has already proven to be a very powerful tool for understanding social interactions (see Schwarz, 1996). Such norms can lead an assertion made by one person to yield surprising changes in others' beliefs (Gruenfeld & Wyer, 1992). Such norms can shape the inferences an observer makes about the beliefs of a speaker (Swann, Giuliano, & Wegner, 1982). And such norms can lead people to overuse or underuse certain types of information in making decisions (Krosnick et al., 1990; Schwarz et al., 1991; Tetlock et al., 1996). In all of these cases, effects occurred because people presumed that speakers were conforming to conversational norms. Our results add to this literature by documenting what can occur when speakers violate conversational conventions.

In doing so, our findings place the notion of conversational conventions in direct relevance to a larger concern about the validity of data collected via closed-ended, structured questionnaires. Suchman and Jordan (1992) pointed out that conventional questionnaires script interpersonal exchanges that are quite different from normal, everyday conversations. Unlike in such conversations, respondents cannot control what topics are discussed, are limited in the language they can use to answer questions, are generally prohibited from expressing nuances and hesitations, and usually cannot negotiate meanings or effectively correct misunderstandings. Suchman and Jordan (1992) argued that such discrepancies between ordinary communication and questionnaire-based measurement compromise the accuracy of data collected by the latter. However, surprisingly few studies have been conducted to assess whether making questionnaire-based data collection more like ordinary conversation would actually increase the validity of the obtained data (though see Schober & Conrad, 1997).

We have documented another way in which questionnaires can conform to or depart from conventions of everyday conversation. And we have shown that departures from those conventions do indeed compromise data quality. There-

fore, it seems that questionnaire designers could benefit from gaining further systematic insight into all the conversational conventions potentially applicable to the sorts of questions they typically ask and from attempting to conform to those conventions whenever possible. And for psychologists, learning about such conversational conventions can provide new insights into the dynamics of social interactions more generally.

REFERENCES

- Butt, J. (1996). *Spanish grammar*. New York: Oxford Univ. Press.
- Carp, F. M. (1974). Position effects on interview responses. *Journal of Gerontology*, **29**, 581–587.
- Carson, R. T., & Mitchell, R. C. (1995). Sequencing and nesting in contingent valuation surveys. *Journal of Environmental Economics and Management*, **28**, 155–173.
- Carvajal, H., & Pauls, K. K. (1995). Relationships among Graduate Record Examination scores, Wechsler Adult Intelligence Scale–Revised IQs and undergraduate grade point average. *College Student Journal*, **29**, 414–416.
- Ceci, S. J. (1991). How much does schooling influence general intelligence and its cognitive components?: A reassessment of the evidence. *Developmental Psychology*, **27**, 703–722.
- Clark, H. H. (1992). *Arenas of language use*. Chicago, IL: Univ. of Chicago Press.
- Clark, H. H. (1996). *Using language*. Cambridge, UK: Cambridge Univ. Press.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.
- College Entrance Examination Board. (1997, June). *Research notes: Common sense about SAT score differences and test validity* (RN-01). New York: Author.
- Coney, K. A. (1977). Order-bias: The special case of letter preference. *Public Opinion Quarterly*, **41**, 385–388.
- Cooper, W. E., & Ross, J. R. (1975). World order. In R. E. Grossman, L. J. San, & T. J. Vance (Eds.), *Paper from the parasession on functionalism*. Chicago: Chicago Linguistic Society.
- Dulany, D. E., & Hilton, D. J. (1991). Conversational implicature, conscious representation, and the conjunction fallacy. *Social Cognition*, **9**, 85–110.
- Fazio, R. H. (1990). A practical guide to the use of response latency in social psychological research. In C. Hendrick & M. S. Clark (Eds.), *Research methods in personality and social psychology: Review of personality and social psychology* (Vol. 11, pp. 74–97). Newbury Park, CA: Sage.
- Garcia Rojas, M. A., & Molesworth, D. J. (1996). *Unraveling spanglish: A practical guide to language interference*. Tunja: Universidad Pedagogica y Tecnologica de Columbia.
- Grandgent, C. H., & Wilkins, E. H. (1915). *Italian grammar*. New York: Heath.
- Grice, H. P. (1975). Logic and conversation. In P. Cole (Ed.), *Syntax and semantics, 9: Pragmatics* (pp. 113–128). New York: Academic Press.
- Gruenfeld, D. H., & Wyer, R. S. (1992). Semantics and pragmatics of social influence: How affirmations and denials affect beliefs in referent propositions. *Journal of Personality and Social Psychology*, **62**, 38–49.
- Guilford, J. P. (1954). *Psychometric methods*. New York: McGraw–Hill.
- Hammer, A. E., & Durrell, M. (1991). *Hammer's German grammar and usage*. New York: Arnold.
- Hawkins, R. D., & Towell, R. (1996). *French grammar and usage*. London: Arnold.
- Hoch, S. J. (1984). Availability and interference in predictive judgment. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **10**, 649–662.
- Jacobs, S. S. (1972). A validity study of the acquiescence scale of the Holland Vocational Preference Inventory. *Educational and Psychological Measurement*, **32**, 477–480.
- Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, **80**, 237–251.
- King, G. (1988). Statistical models for political science event counts: Bias in conventional procedures and evidence for the exponential Poisson regression model. *American Journal of Political Science*, **32**, 838–863.

- Koriat, A., Lichtenstein, S., & Fischhoff, B. (1980). Reasons for confidence. *Journal of Experimental Psychology: Human Learning and Memory*, **6**, 107–118.
- Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*, **5**, 213–236.
- Krosnick, J. A. (1992). The impact of cognitive sophistication and attitude importance on response order effects and question order effects. In N. Schwarz & S. Sudman (Eds.), *Order effects in social and psychological research* (pp. 203–218). New York: Springer-Verlag.
- Krosnick, J. A. (1993). Maximizing questionnaire quality. In J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), *Measures of political attitudes* (pp. 37–57). New York: Academic Press.
- Krosnick, J. A., & Fabrigar, L. R. (in press). *Designing great questionnaires: Exploring the psychology of question-answering*. New York: Oxford Univ. Press.
- Krosnick, J. A., Li, F., & Lehman, D. R. (1990). Conversational conventions, order of information acquisition, and the effect of base rates and individuating information on social judgments. *Journal of Personality and Social Psychology*, **59**, 1140–1154.
- Lawlor, S., Richman, C. L., & Richman, S. (1997). The validity of using the SAT as a criterion for black and white students' admission to college. *College Student Journal*, **31**, 507–515.
- Levinson, S. C. (1983). *Pragmatics*. Cambridge, UK: Cambridge Univ. Press.
- Matthews, C. O. (1927). The effect of position of printed response words upon children's answers to questions in two-response types of tests. *Journal of Educational Psychology*, **18**, 445–457.
- Moore, D. W., & Newport, F. (1996). *Public policy questions and response order: Prevalence of the recency effect*. Paper presented at the annual meeting of the American Association for Public Opinion Research, St. Louis, MO.
- Mosier, C. I., & Price, H. G. (1945). The arrangement of choices in multiple choice questions and a scheme for randomizing choices. *Educational and Psychological Measurement*, **5**, 379–382.
- Narayan, S., & Krosnick, J. A. (1996). Education moderates some response effects in attitude measurement. *Public Opinion Quarterly*, **60**, 58–88.
- Nisbett, R. E., Zukier, H., & Lemley, R. E. (1981). The dilution effect: Non-diagnostic information weakens the implications of diagnostic information. *Cognitive Psychology*, **13**, 248–277.
- Oppenheim, A. N. (1992). *Questionnaire design, interviewing and attitude measurement*. New York: Pinter.
- Rea, L. M., & Parker, R. A. (1997). *Designing and conducting survey research*. San Francisco: Jossey-Bass.
- Rosenthal, R., & Rubin, D. B. (1986). Meta-analytic procedures for combining studies with multiple effect sizes. *Psychological Bulletin*, **99**, 400–406.
- Schober, M. F., & Conrad, F. G. (1997). Does conversational interviewing reduce survey measurement error? *Public Opinion Quarterly*, **61**, 576–602.
- Schuman, H., & Presser, S. (1981). *Questions and answers in attitude surveys*. New York: Academic Press.
- Schwarz, N. (1996). *Cognition and communication: Judgmental biases, research methods, and the logic of conversation*. Mahwah, NJ: Erlbaum.
- Schwarz, N., Strack, F., Hilton, D. J., & Naderer, G. (1991). Base rates, representativeness, and the logic of conversation: The contextual relevance of 'irrelevant information.' *Social Cognition*, **9**, 67–84.
- Shank, R. C., & Abelson, R. P. (1977). *Scripts, plans, goals, and understanding: An inquiry into human knowledge structure*. Hillsdale, NJ: Erlbaum.
- Snyder, M., & Gangestad, S. (1981). Hypothesis-testing processes. In J. H. Harvey, W. Ickes, & R. F. Kidd (Eds.), *New directions in attribution research* (Vol. 3). Hillsdale, NJ: Erlbaum.
- Strack, F., & Schwarz, N. (1992). Communicative influences in standardized question situations: The case of implicit collaboration. In G. R. Semin & K. Fiedler (Eds.), *Language, interaction, and social cognition* (pp. 173–193). London: Sage.
- Strack, F., Schwarz, N., & Wänke, M. (1991). Semantic and pragmatic aspects of context effects in social and psychological research. *Social Cognition*, **9**, 111–125.

- Suchman, L., & Jordan, B. (1992). Validity and the collaborative construction of meaning in face-to-face surveys. In J. M. Tanur (Ed.), *Questions about questions: Inquiries into the cognitive bases of surveys*. New York: Sage.
- Swann, W. B., Guiliano, T., & Wegner, D. M. (1982). Where leading questions can lead: The power of conjecture in social interaction. *Journal of Personality and Social Psychology*, **42**, 1025–1035.
- Tetlock, P. E., Lerner, J. S., & Boettger, R. (1996). The dilution effect: Judgmental bias, conversational convention, or a bit of both? *European Journal of Social Psychology*, **26**, 915–934.
- Tourangeau, R., & Rasinski, K. A. (1988). Cognitive processes underlying context effects in measurement. *Psychological Bulletin*, **103**, 299–314.
- Wason P. C., & Johnson-Laird, P. N. (1972). *Psychology of reasoning: structure and content*. Cambridge, MA: Harvard Univ. Press.