Comparisons of Party Identification and Policy Preferences: The Impact of Survey Question Format

Jon A. Krosnick; Matthew K. Berent

American Journal of Political Science, Vol. 37, No. 3 (Aug., 1993), 941-964.

Stable URL:
http://links.jstor.org/sici?sici=0092-5853%28199308%2937%3A3%3C941%3ACOPIAP%3E2.0.CO%3B2-L

American Journal of Political Science is currently published by Midwest Political Science Association.

*Comparisons of Party Identification and Policy Preferences: The Impact of Survey Question Format**

Jon A. Krosnick, *Department of Psychology, Ohio State University*
Matthew K. Berent, *Department of Psychology, Ohio State University*

Much research has suggested that citizens' political party affiliations are more persistent over time and more psychologically consequential than are their attitudes toward government policies. However, most surveys have measured party identification with branching questions in which all response alternatives were verbally labeled, whereas policy preferences have typically been measured using nonbranching questions with incomplete verbal labeling of response alternatives. We report eight experiments, involving telephone interviewing, face-to-face interviewing, and self-administered questionnaires, demonstrating that fully labeled branching measures of party identification and policy attitudes are more reliable than partially labeled nonbranching measures of those attitudes. This difference seems to be attributable to effects of both verbal labeling and branching. Thus, it appears that previous findings regarding differences between party identification and policy preferences are partly due to the failure to equate the formats of survey questions measuring those attitudes.

Introduction

According to political theorists, democratic governments can maintain stability and legitimacy by implementing policies favored by majorities of their citizens (see Dahl 1956, 1989; Pennock 1979). This is presumed to occur if citizens elect representatives who share their preferences regarding government policy. Such a vision of democracy requires that citizens attend closely to the policymaking activities of government, formulate well-informed preferences about what they would like government to do on specific issues, and use these preferences to make vote choices in elections.

*Portions of this paper were presented at the annual meetings of the American Association for Public Opinion Research, the International Society for Political Psychology, the Midwestern Psychological Association, and the American Political Science Association. The authors wish to thank Peggy Caldwell, Schelli Cavenaugh, Tom Herby, and Konnie Toth for conducting the interviewing for Study 1, Sondra Crayton and Regina Owens for conducting the interviewing for Study 5, and Santa Traugott for all her help coordinating our sections of the National Election Studies. The authors also wish to thank Paul Beck, David Boninger, George Bishop, Stanley Presser, Vincent Price, the National Election Study Board of Overseers, and especially Donald Kinder and John Zaller for helpful comments and suggestions.

American Journal of Political Science, Vol. 37, No. 3, August 1993, Pp. 941–964

The viability of any specific democratic government can therefore be assessed partly by the degree to which its citizens behave in this fashion. And since the earliest quantitative analyses of sample surveys in the United States, investigators have asserted that the U.S. citizenry falls far short of the mark. For example, in *The American Voter,* Campbell et al. (1960) concluded that policy preferences are only peripherally involved in most Americans' political thinking. Many respondents in their surveys lacked opinions on policy matters, were unaware of current government policy on issues, and saw no differences between the political parties in terms of policy stands. And the opinions people did offer seemed to be unanchored by firm beliefs or accurate knowledge about political affairs (see also Converse 1964). Later studies have replicated these findings and demonstrated as well that policy preferences explain only very little variance in people's vote choices (e.g., Nie, Verba, and Petrocik 1979).

Instead of policy issues, what has seemed to reside at the center of most Americans' political thinking is their identification with political parties (Campbell et al. 1960; Campbell, Gurin, and Miller 1954; Converse 1966). According to *The American Voter,* most citizens in the 1950s inherited a party affiliation from their parents, carried it intact throughout their lives, and used it extensively to evaluate political events and to make vote choices. Certainly, the strength and role of party identification has declined since then (see Kinder and Sears 1985, 684–86; Wattenberg 1984), and we have come to recognize that party identification is in fact responsive to short-term political events to some degree instead of being permanently crystallized during childhood (e.g., Fiorina 1981; Franklin 1983). Nonetheless, Kinder and Sears pronounced in 1985 that "party identification remains the single most important determinant of individual voting decisions" (686) and presumably of political judgments and behavior more generally.

A cornerstone of the evidence reinforcing this conclusion involves attitude stability. Americans' reports of their party affiliations are a great deal more consistent over time than are their reports of preferences on policy issues. This has been shown using the 1956–60 National Election Panel Study (Converse 1964), the 1972–76 NES Panel (Converse and Markus 1979; Sears 1983), the 1980 NES Panel (Markus 1982), Jennings and Niemi's Youth-Parent Socialization Panel Study (Jennings and Markus 1984; Jennings and Niemi 1981), and Patterson's 1976 Election Panel Study (Feldman 1989). On the basis of this pattern, many analysts have concluded that "partisan loyalties are more stable over time than any other political attitude" (Abramson 1983, 99; see also Sears 1983).

In this paper, we explore a possible alternative interpretation for this

finding. In previous studies (Converse 1964; Converse and Markus 1979; Jennings and Markus 1984; Jennings and Niemi 1981; Markus 1982; Sears 1983), the stability of party identification and policy preferences has been assessed by examining zero-order test-retest correlations. But such correlations reflect both the amount of attitude change that takes place during a given time period and the amount of random measurement error in reports of the attitude (see, e.g., Alwin 1973). Therefore, evidence that reports of party identification are more consistent over time than reports of policy attitudes may be due to differences in either or both of these processes.

In fact, there are two reasons why the NES party identification measures may be more reliable than the policy attitude measures in those surveys.[1] First, party identification has been measured via a series of branching questions in which respondents are first asked whether they consider themselves to be Republicans, Democrats, or independents. People who report identifying with a party are then asked whether they do so strongly or weakly, and self-proclaimed independents are asked whether they lean toward one party or the other. Thus, respondents are led through a two-step process in which they first report attitude direction and then report attitude intensity.

In contrast, the 1950s, 1970s, and 1980s NES policy attitude measures involved one-step procedures in which respondents placed themselves directly on bipolar attitude scales, thus indicating attitude direction and extremity simultaneously. Armstrong, Denniston, and Gordon (1975) have shown that decomposing a complex decision task into simpler component decisions increases the precision of the final result. It may therefore be that decomposing attitude reports through branching increases their reliability. This could partly explain why party identification reports are more consistent over time than policy attitude reports in the NESs.

A second possible explanation for this difference involves the extent of verbal labeling of response alternatives. In the 1970s and 1980s NESs, the policy attitude measures asked respondents to place themselves on seven-point scales on which only the endpoints were labeled with words. In contrast, the NES party identification questions asked respondents to select among answer choices that were all verbally labeled. A number of experimental studies have shown that verbal labeling of response alternatives enhances the reliability of attitude reports, presumably because it

[1]Patterson's Election Panel Study (Feldman 1989) used NES questions to measure party identification and policy preferences. Jennings and Niemi's (1981) study used NES party identification measures, though their policy attitude measures were somewhat different.

clarifies the meanings of otherwise ambiguous scale points (Madden 1960; Peters and McCormick 1966; Zaller 1988; cf. Finn 1972). Therefore, the difference in verbal labeling could also partly account for the reduced over-time consistency of policy attitude reports as compared with reports of party identification.

Some indirect support for these hypotheses has been provided by Feldman (1989) and Krosnick (1991b), who showed that the NES party identification measure is more reliable than NES policy attitude measures. More directly supportive is evidence reported by Aldrich et al. (1982). Using the 1979 National Election Pilot Study data, these investigators compared partially labeled nonbranching measures of three policy attitudes with fully labeled branching versions of the same questions. Where differences appeared between the two formats, the latter questions were more strongly related to other political attitudes and produced stronger test-retest associations.

However, there are three primary problems with Aldrich et al.'s (1982) evidence. First, two of the fully labeled branching questions were asked only during the first of two interviews, while the partially labeled nonbranching versions of those questions were asked only during the second interview. Thus, question format and time of measurement were confounded. Second, instead of reflecting enhanced reliability, Aldrich et al.'s results may have occurred because the fully labeled branching questions produced more extreme responses (see Aldrich et al. 1982, 401–02), which increases response variance and thereby strengthens standardized measures of association such as those that Aldrich et al. examined (see Duncan 1975). Finally, Aldrich et al. (1982) did not subject any of the differences between question formats that they observed to statistical significance tests, so it is difficult to assess their robustness.

We set out to determine more definitively whether branching and verbal labeling of response alternatives affect the reliability of the NES measures of party identification and policy preferences. To do so, we conducted eight experiments that are described below. The studies differed from one another in terms of a variety of methodological characteristics, such as subject population, mode of data collection, question formats examined, and attitudes measured (see Table 1).

Study 1

Data

For our first study, a random digit dial (RDD) telephone survey was conducted with 69 residents of the Columbus, Ohio, metropolitan area. Approximately one month later, 63 of these individuals were successfully

Table 1. Methodological Characteristics of Studies

Study	Respondents	Sample Size[a]	Data Collection Mode	Question Forms			Attitudes			Average Overall Effect Size (r)
				Partially Labeled Nonbranching	Fully Labeled Nonbranching	Fully Labeled Branching	Party ID	Ideology	Policy Issues	
1.	General population	63	telephone	√		√	√	√	√	.24
2.	General population	494	telephone	√		√	√	√	√	.11
3.	College students	77	self-administered	√		√	√	√	√	.16
4.	General population	2,004	face to face	√		√			√	.06
5.	College students	106	face to face	√	√	√	√	√	√	.11[b]
6.	General population	1,385	telephone	√	√	√			√	.08[c]
7.	General population	274	telephone		√	√		√		.17
8.	General population	998	telephone		√	√		√		.06

[a]Sample sizes are the numbers of individuals who were asked all of our experimental questions.

[b]This figure is based on comparisons of the fully labeled nonbranching questions with the fully labeled branching questions.

[c]This figure is based on comparisons of the partially labeled nonbranching question with the fully labeled nonbranching question.

recontacted for a second interview. Four trained interviewers were randomly assigned to telephone numbers for initial and follow-up interviews.

All respondents were asked about their political party identifications, their ideological orientations, and their attitudes toward federal aid to minority groups, defense spending, and U.S. involvement in Central America. Each respondent was randomly assigned to receive one of two versions of these questions during both interviews: fully labeled branching or partially labeled nonbranching.

The fully labeled branching questions used the format of the NES party identification question, and the partially labeled nonbranching questions used the format of the NES policy attitude questions. For example, here are the two versions of the defense spending question we examined:

> *Partially labeled nonbranching.* "There has been a lot of debate recently about defense spending. Some people believe that the U.S. should spend much less money for defense. Suppose these people are at one end of a seven-point scale, at point number 1. Others feel that defense spending should be greatly increased. Suppose these people are at the other end of the scale—at point number 7. And, of course, other people have opinions somewhere in between, at points 2, 3, 4, 5, and 6. Where would you place yourself on this scale?"
>
> *Fully labeled branching.* "There has been a lot of debate recently about defense spending. Do you think the U.S. should spend less money on defense, more money on defense, or continue spending about the same amount on defense?" [If less:] "Would you say we should spend a lot less, somewhat less, or a little less?" [If more:] "Would you say we should spend a lot more, somewhat more, or a little more?"

Answers to the fully labeled branching questions were transformed into a seven-point scale as follows: "a lot less" = 1; "somewhat less" = 2; "a little less" = 3; "the same" = 4; "a little more" = 5; "somewhat more" = 6; and "a lot more" = 7.

Results

The top panel of Table 2 displays the percentages of respondents who provided the same answer to each question during both interviews. The larger this percentage, the greater consistency there is in responses, and the more reliable the measure presumably is.[2] As expected, the com-

[2]Other statistics assessing the test-retest consistency of the attitude reports, including Gamma, Tau-*b*, Pearson product-moment correlations, and unstandardized regression coef-

bined reliability of the five fully labeled branching questions was substantially and significantly greater than the combined reliability of the five partially labeled nonbranching questions.[3]

Study 2

Given the encouraging initial results of Study 1, we attempted to replicate them using data from the 1988 National Election Study and the 1989 National Election Pilot Study.

Data

For the 1988 NES, a representative national sample of 1,640 U.S. adults were interviewed face to face in their homes twice, first between 6 September and 7 November 1988 and then between 18 November 1988 and 30 January 1989. A stratified random sample of 855 of these individuals were selected to be reinterviewed for the 1989 NES Pilot. Of these people, 614 were successfully reinterviewed by telephone for the first wave during July and August 1989, and 494 of them were successfully interviewed again by telephone during September and October 1989.

All respondents answered questions during both 1989 interviews about their party identifications, ideological orientations, and attitudes on defense spending, U.S. involvement in Central America, and gun control. Each respondent was randomly assigned to receive one of four forms of these questions during both interviews. Two question forms were partially labeled nonbranching, and two were fully labeled branching. One of the partially labeled nonbranching sets included a no-opinion filter in each question, and the other did not. Similarly, one fully labeled branching question set included no-opinion filters, and the other did not.[4]

---

ficients, all supported the same conclusions reported in the text for this experiment and all experiments to follow.

All of these statistics actually reflect both measurement unreliability and the amount of true attitude change that took place during the one-month time period. Therefore, our estimates of reliability underestimate each item's true reliability if some attitude change occurred. However, if attitude change did occur, it most likely occurred in equal amounts in the two groups of respondents because respondents were randomly assigned to these groups. Therefore, *differences* between the groups in terms of the reliability estimates here and in the studies that follow are almost certainly not distorted by the presence of attitude change.

[3]All significance levels reported in this paper for tests of our primary hypothesis are one-tailed because we made strong a priori predictions.

[4]The ideology measure included a no-opinion filter in all of the question sets, and the wording of the verbal labels was varied instead. In two question sets, the most extreme response alternatives were labeled "very" liberal or conservative, whereas in the other two question sets, they were labeled "extremely" liberal or conservative.

Table 2. Consistency of Attitude Reports over 1–3 Months

Attitude	Question Form: Partially Labeled Nonbranching %	Question Form: Fully Labeled Nonbranching %	Question Form: Fully Labeled Branching %	Difference %	Significance
Study 1:					
Party identification	29.6	69.4		39.8	$p < .001$
Ideological orientation	48.1	51.4		3.3	ns
Aid to minorities	37.0	64.7		27.7	$p < .016$
Defense spending	51.9	83.3		31.4	$p < .004$
Central America	29.6	50.0		20.4	$p < .055$
Combined	39.3	63.8		24.5	$p < .00001$
	(27)[a]	(36)			
Study 2:					
Party identification	59.4	67.2		7.8	$p < .038$
Ideological orientation	42.1	51.7		9.6	$p < .029$
Defense spending	40.4	51.0		10.6	$p < .014$
Central America	32.2	48.8		16.6	$p < .0002$
Gun control	45.6	62.1		16.5	$p < .0002$
Combined	44.3	56.6		12.3	$p < .00001$
	(226)	(217)			

Study 3:					
Party identification	43.9		84.6	40.7	$p < .001$
Ideological orientation	58.5		53.8	−4.7	ns
Defense spending	43.9		64.1	20.2	$p < .036$
Central America	41.5		53.8	12.3	ns
Aid to the poor	43.9		55.3	11.4	ns
Combined	46.3		62.4	16.1	$p < .0007$
	(41)		(36)		
Study 5:					
Party identification		77.8	94.3	16.5	$p < .023$
Ideological orientation		52.8	68.6	15.8	$p < .09$
Defense spending		58.3	60.0	1.7	ns
Central America		50.0	45.7	−4.3	ns
Social programs		50.0	73.5	23.5	$p < .022$
Combined		57.8	68.4	10.6	$p < .02$
		(36)	(35)		

Note: Table entries are the proportions of respondents who provided identical responses across two interviews.

[a]Numbers in parentheses are numbers of cases.

Results

Reliability over two months. Replicating Study 1, the combined consistency of responses across the two 1989 interviews for the fully labeled branching items was again significantly larger than the combined consistency for the partially labeled nonbranching items (see the second panel of Table 2).[5] Consistent with Krosnick's (1991a) expectation, labeling and branching were less helpful among respondents with some education beyond high school ($\Delta b = 8.4\%$, $\chi^2(1) = 7.81$, $p < .003$) than among respondents without a high school diploma ($\Delta b = 12.6\%$, $\chi^2(1) = 4.94$, $p < .013$) or among respondents with a high school diploma but no additional education ($\Delta b = 17.2\%$, $\chi^2(1) = 22.51$, $p < .0001$).

Reliability over 8–10 months. During the 1988 preelection interview, all respondents were asked questions measuring four of our target attitudes. If the 1989 fully labeled branching questions were more reliable than their partially labeled nonbranching counterparts, then the former should be associated more strongly than the latter with 1988 reports of the same attitudes. And as the top panel of Table 3 displays, unstandardized regression coefficients estimating these associations confirm that expectation.

Associations with other political judgments. In order to further test our hypothesis, we examined associations between our 1989 target attitude items and other political judgments assessed in 1988 and 1989. As expected, unstandardized regression coefficients combined across the two 1989 waves and the five issues showed that our fully labeled branching questions were more strongly associated with 1989 evaluations of President Reagan's job performance than our partially labeled nonbranching questions were (see the top panel of Table 4). Similarly, an unstandardized regression coefficient assessing the association between 1989 policy attitudes and 1988 reports of vote choice (Bush or Dukakis), combined across the two 1989 waves and the five issues, was again significantly larger for the fully labeled branching questions than for the partially labeled nonbranching questions (see Table 4).[6]

[5]The labeling and branching effect for party identification and policy preferences was equally strong when a no-opinion filter was included and when it was not. Similarly, the labeling and branching effect on the reliability of ideology reports did not vary depending upon whether the endpoints of the ideology scale were labeled "extremely" or "very."

[6]We replicated these findings regarding associations between policy attitudes and other political judgments using data from the 1982 National Election Study Methods Comparison Project. In that survey, respondents were asked either partially labeled nonbranching or fully labeled branching questions about government aid to minorities, government spending for social welfare programs, defense spending, guaranteed employment, and women's

Table 3. Consistency of Attitude Reports over 8–12 Months

Attitude	Question Form: Partially Labeled Nonbranching	Question Form: Fully Labeled Nonbranching	Question Form: Fully Labeled Branching	Difference	Significance
Study 2:[a]					
Party identification	.61 (311)[b]		.86 (300)	.25	$p < .0001$
Ideological orientation	.57 (202)		.60 (197)	.03	ns
Defense spending	.53 (262)		.75 (252)	.22	$p < .001$
Central America	.92 (269)		1.07 (253)	.15	ns
Study 6:					
Social welfare	.38 (324)	.52 (348)		.14	$p < .03$
Study 7:					
Ideology		.56 (84)	.89 (96)	.33	$p < .011$

Note: Table entries are unstandardized regression coefficients.

[a]These associations were computed using data from the first wave of the 1989 survey. When they were computed using the second wave data instead, the same pattern of results was obtained.

[b]Numbers in parentheses are number of cases.

Table 4. Consistency between Target Attitudes and Criterion Items

Attitude	Question Form: Partially Labeled Nonbranching	Question Form: Fully Labeled Nonbranching	Question Form: Fully Labeled Branching	Difference	Significance
Study 2:					
Five attitudes with					
Presidential performance	.35		.46	.11	$p < .005$
	(255)[a]		(237)		
Vote choice	.91		1.10	.19	$p < .02$
	(242)		(220)		
Study 4:					
South Africa attitudes with					
Race	1.11		1.84	.73	$p < .002$
	(667)		(598)		
Attitude toward blacks	1.76		2.38	.62	$p < .09$
	(653)		(582)		

Study 6:					
Social welfare attitudes with party identification	.15 (347)	.29 (378)		.14	$p < .008$
Study 7:					
Ideology with presidential performance		.21 (101)	.43 (122)	.22	$p < .03$
Study 8:					
Ideology with candidate Preference		1.43 (336)	1.93 (331)	.50	$p < .04$
Party identification		.25 (333)	.31 (329)	.06	$p < .14$

Note: Table entries are unstandardized regression coefficients.

[a]Numbers of cases are in parentheses.

Differences between attitude objects. As in analyses of prior NES panel studies (e.g., Krosnick 1991b), the reliability of the traditional NES party identification measure in Table 2 (67.2%) was larger than the reliability of the traditional NES policy attitude measures addressing defense spending (40.4%, $\chi^2(1) = 34.01$, $p < .0001$), Central America (32.2%, $\chi^2(1) = 58.72$, $p < .0001$), and gun control (45.6%, $\chi^2(1) = 22.98$, $p < .0001$). If fully labeled branching questions had been used to assess both party identification and policy attitudes in prior studies, would this difference in reliability have disappeared?

According to these data, the answer is no: controlling for question format did reduce the difference in reliability between measures of policy attitudes and party identification, but it was not completely eliminated. In the case of the partially labeled nonbranching items, party identification (59.4%) was significantly more reliable than defense spending (40.4%, $\chi^2(1) = 17.16$, $p < .0001$), Central America (32.2%, $\chi^2(1) = 36.04$, $p < .0001$), gun control (45.6%, $\chi^2(1) = 9.40$, $p < .002$), and the combination of the three policy attitudes (39.5%, $\chi^2(1) = 29.41$, $p < .0001$). In the case of the fully labeled branching questions, party identification (67.2%) was also significantly more reliable than defense spending (51.0%, $\chi^2(1) = 12.37$, $p < .0002$), Central America (48.8%, $\chi^2(1) = 15.59$, $p < .0001$), and the combination of the three issues (54.1%, $\chi^2(1) = 12.48$, $p < .0002$), although not gun control (62.1%, $\chi^2(1) = 1.36$, $p < .13$) (see Krosnick 1991b for a discussion of possible sources of these remaining differences).

Administration time. During the 1989 interviews, four of our target questions (party identification, defense spending, Central America, and gun control) were all measured in one minute and 43 seconds on average with the fully labeled branching questions, whereas it took two minutes and 13 seconds on average to measure all of these attitudes using the partially labeled nonbranching questions ($\chi^2(6) = 60.16$, $p < .00001$). Thus, a sequence of two fully labeled branching questions can measure an attitude more quickly than a single partially labeled nonbranching question, saving approximately seven seconds per attitude (or 23%). This is presumably so because a partially labeled nonbranching question usually entails reading about twice as many words as are involved in a sequence of two fully labeled branching questions.

---

rights. The branching versions of the questions were notably more strongly associated with candidate preferences, party identification, and other global political judgments. We do not report these results in detail here because they simply mirror Study 2's results using equivalent data.

Study 3

Given this support for our hypotheses from telephone surveys, we next explored whether the labeling-branching effect could be observed using self-administered questionnaires.

Data

Seventy-seven undergraduates at Ohio State University participated in this study in partial fulfillment of an introductory psychology course requirement. Respondents made two visits to our laboratory (separated by approximately one month), during which they completed questionnaires in individual cubicles. At both waves, respondents reported their party identifications, ideological orientations, and attitudes on defense spending, U.S. involvement in Central America, and aid to the poor. Each respondent was again randomly assigned to receive either partially labeled nonbranching or fully labeled branching questions during both visits.

Results

As expected, verbal labeling and branching markedly increased the reliability of the attitude reports for all five items combined (see the third panel of Table 2).

Study 4

Our next study assessed the generality of the verbal labeling and branching effect to yet another data collection mode: face-to-face interviewing. To do so, we analyzed data from the 1990 National Election Study.

Data

For the 1990 NES, a representative national sample of 2,004 U.S. adults were interviewed in their homes between 7 November 1990 and 26 January 1991. All respondents were asked a question measuring their attitudes toward sanctions against South Africa in either a partially labeled nonbranching form or a fully labeled branching form (determined randomly). No-opinion filters were included in both forms of the question, and show cards displayed the answer choices for the partially labeled nonbranching question only.

Results

In order to assess the reliability of the target item, we examined associations between it and two criterion items: respondents' race (coded

dichotomously: white vs. nonwhite), and attitudes toward blacks (as measured by a 101-point feeling thermometer).[7] As expected, race was more strongly associated with the fully labeled branching South Africa question than with the partially labeled nonbranching version (see the second panel of Table 4). Also as expected, this difference was larger among respondents without a high school diploma ($\Delta b = .92$, $z = 1.66$, $p < .05$) than among respondents with a high school diploma but no additional education ($\Delta b = .77$, $z = 1.79$, $p < .04$) and respondents with some education beyond high school ($\Delta b = .54$, $z = 1.39$, $p < .09$).

Similarly, attitudes toward blacks were more strongly associated with the fully labeled branching South Africa question than with the partially labeled nonbranching version (see Table 4). This difference was again larger among respondents without a high school diploma ($\Delta b = .87$, $z = 0.86$, $p < .20$) than among respondents with a high school diploma but no additional education ($\Delta b = .39$, $z = 0.52$, ns) and respondents with some education beyond high school ($\Delta b = .56$, $z = 0.83$, ns).

Study 5

Our fifth study explored whether the labeling and branching manipulation enhanced reliability because of the verbal labeling per se, the branching per se, or both. We also used a design that overcame a confound in Study 4: show cards were used to display the answer choices for the partially labeled nonbranching questions only. Most likely, this would have increased the reliability of these questions, thus working against our hypothesis. In Study 5, we overcame this problem by displaying all answer choices on show cards.

Data

A total of 106 undergraduates at Ohio State University participated in two identical face-to-face interviews (approximately one month apart) in partial fulfillment of an introductory psychology course requirement. Two trained interviewers were randomly assigned to respondents.

During both interviews, respondents answered questions about their party identifications, ideological orientations, attitudes on defense spending, U.S. involvement in Central America, and government spending for

[7]In order to serve as a criterion item, a question had to be correlated with the target item in the sample as a whole and especially among respondents low in education (among whom the labeling and branching effect should be greatest; see Krosnick 1991a).

A labeling and branching experiment was included in the 1990 NES involving a measure of attitudes toward limiting foreign imports to the United States, but we were unable to find any suitable criterion items for assessing the reliability of this question.

social programs. Each respondent was randomly assigned to receive one of three question formats during both interviews: partially labeled nonbranching, fully labeled branching, or fully labeled nonbranching (i.e., seven-point scales with verbal labels on every alternative). All respondents were handed a show card displaying the response alternatives each time they were asked a question.

Results

We again assessed reliability using the proportion of respondents who provided identical answers to the same question during both interviews. To our surprise, the combined reliability of the five partially labeled nonbranching items (58.9%) was not significantly different from the combined reliability of the fully labeled nonbranching items (57.8%, $\chi^2(1) = 0.04$, ns), and none of the differences between reliabilities for the individual items was statistically significant. Thus, verbal labeling alone apparently had no effect on reliability. However, the combined reliability estimate for the fully labeled branching format was significantly greater than that for the fully labeled nonbranching format (see the bottom panel of Table 2). Thus, branching did seem to enhance reliability.

Study 6

To further explore these issues with a more heterogeneous national sample of adults, we analyzed data from the 1990 National Election Study and the 1991 NES Pilot Study.

Data

Of the 2,004 U.S. adults interviewed for the 1990 NES between 7 November 1990 and 26 January 1991, 1,385 were reinterviewed by telephone during June and July 1991 for the 1991 NES Pilot Study. During the 1990 interview, all respondents were asked the standard NES partially labeled nonbranching question about increasing or decreasing government spending for social welfare programs. During their 1991 interviews, respondents were again asked about this issue, but in one of three different ways (determined randomly): partially labeled nonbranching, fully labeled nonbranching, or fully labeled branching. A no-opinion filter was included in all three versions.

Results

Answers to the partially labeled nonbranching question were quite a bit less consistent with 1990 reports of the same attitude than were answers to the fully labeled nonbranching question (see the middle panel of Table 3). However, answers to this latter question were no less consistent

with previous attitude reports than were answers to the fully labeled branching question ($b = .47$, $n = 333$; $z = 0.05$, ns). Thus, it appears that verbal labeling increased reliability but branching did not. As expected, the effect of labeling was larger among low education respondents ($\Delta b = .24$, $z = 1.16$, $p < .13$) and moderate education respondents ($\Delta b = .26$, $z = 1.91$, $p < .03$) than among high education respondents ($\Delta b = .04$, $z = 0.41$, ns).

Similarly, answers to the partially labeled nonbranching question were less strongly associated with 1990 reports of party identification than were answers to the fully labeled nonbranching question (see the third panel of Table 4). But answers to this latter question were no less consistent with party identification than answers to the fully labeled branching question ($b = .33$, $n = 354$; $z = .04$, ns). And again, the labeling-induced increase in the association between 1991 government spending attitudes and 1990 party identification was larger in the low education group ($\Delta b = .23$, $z = 1.40$, $p < .09$) than in the moderate education group ($\Delta b = .10$, $z = 1.01$, $p < .16$) and the high education group ($\Delta b = .14$, $z = 1.79$, $p < .04$).

Study 7

The results of Study 6 appear to conflict with those of Study 5, where we found an effect of branching but no effect of labeling. However, the failure to find a labeling effect in Study 5 is understandable in light of Study 6's finding of an interaction involving education. The respondents in Study 5 were OSU undergraduates, who had high levels of educational attainment. Therefore, because Study 6's results suggest that labeling is of little help to highly educated individuals, it is no surprise that labeling had no impact on the reliability of the attitude reports made by Study 5's respondents.

It is more difficult to understand the failure of Study 6 to replicate the branching effect uncovered in Study 5. One possible explanation is that Study 5 involved face-to-face interviews with show cards, whereas Study 6 involved telephone interviews. We therefore conducted two further studies using telephone interviews in order to test for an effect of branching in general population samples. Our first such attempt involved analysis of data from the 1982 National Election Study and the 1983 NES Pilot.

Data

For the 1982 NES, a representative national sample of 1,418 U.S. adults were interviewed face to face in their homes between 3 November

1982 and 31 January 1983. A stratified random sample of 416 of these individuals was selected for the 1983 NES Pilot, 314 of whom were successfully reinterviewed by telephone during July 1983. Of these individuals, 274 were successfully interviewed again by telephone during August 1983.

During the 1982 interview, all respondents were asked the standard NES ideological orientation question. During their final 1983 interviews, all respondents again answered a question about their ideological orientations in either a fully labeled nonbranching form with a no-opinion filter or a fully labeled branching form without a filter (determined randomly).[8]

Results

As expected, unstandardized regression coefficients indicated that 1983 answers to the branching question were quite a bit more consistent with 1982 ideology reports than were answers to the nonbranching question (see the bottom panel of Table 3). Also, answers to the branching question were more strongly associated with 1983 evaluations of President Reagan's job performance than were answers to the nonbranching question (see the fourth panel of Table 4).

Study 8

We conducted a final study to attempt to replicate this effect with data from the 1982 National Election Study Methods Comparison Project (MCP).

Data

For the 1982 NES MCP, a representative national sample of 998 U.S. adults were interviewed by telephone between 3 November 1982 and 31 January 1983. Respondents were randomly assigned to be interviewed either by the University of Michigan's Survey Research Center or by the University of California at Berkeley's Program in Computer-Assisted Survey Methods. During the interviews, all respondents were asked about their ideological orientations. The Berkeley respondents received a fully labeled nonbranching question, and the Michigan respondents received a fully labeled branching version. Overcoming the confound in Study 7, both question forms included a no-opinion filter.

[8]Omission of the no-opinion filter should presumably decrease the reliability of this item. Given that we expect the branching format to increase item reliability, this omission works against our detecting a reliability increase due to branching.

Results

As expected, the branching question was more strongly associated with candidate preferences (assessed by the difference between respondents' evaluations of Jimmy Carter and Ronald Reagan on 101-point feeling thermometers) than was the nonbranching version (see the bottom panel of Table 4). Similarly, the association between ideology and party identification was stronger for the fully labeled branching question than for the fully labeled nonbranching question, though not quite significantly (see Table 4).

Meta-Analysis

The effects of our labeling and branching manipulations on over-time consistency were clearly statistically significant in each of the studies reported above when we collapsed across all the items involved in each study. However, individual items did not always reveal statistically significant effects on over-time consistency in each test, particularly in the cases of ideology, Central America, and defense spending in Studies 1, 3, and 5. This raises the possibility that although the labeling and branching effect may occur for some attitudes, it may not occur reliably for these attitudes.

The literature on meta-analysis (Rosenthal 1988) has made it clear that even though an effect may not appear significantly in every test of it, the effect may nonetheless be real and robust. In support of this notion here, of the eight marginally significant or nonsignificant effects we observed, six of them are in the expected direction. More important, meta-analyses combining the over-time consistency results of Studies 1, 2, 3, and 5 indicated that the apparently variable effects were indeed robust: the labeling and branching manipulation enhanced reliability for defense spending attitudes ($z = 2.61$, $p = .005$), Central America attitudes ($z = 2.03$, $p = .03$), and ideological orientations ($z = 1.42$, $p = .08$). Furthermore, significant effects were observed on ideology reports in Studies 7 and 8. Thus, the labeling and branching effect appears to be reliable across a range of attitudes.

Overall estimates of effect sizes, averaged across all tests of the branching and labeling manipulations within each study, are displayed in the last column of Table 1 (see Rosenthal 1988). These figures indicate that the manipulations consistently produced modest but reliable improvements in reliability.

Discussion

These studies indicate that decomposing an attitude rating question into fully verbally labeled component questions measuring direction and

extremity enhances the reliability of the obtained attitude reports. We have demonstrated this effect with telephone interviews, face-to-face interviews, and self-administered questionnaires, with representative national samples of adults, a sample of adult residents of Columbus, Ohio, and samples of college undergraduates and for a range of political attitudes, including party identification, ideological orientation, and various policy attitudes. Therefore, the effect seems to be quite robust.

The results of Studies 5–8 suggest that verbal labeling and branching both improve attitude measure reliability. Our demonstration of a verbal labeling effect is consistent with various prior studies (Madden 1960; Peters and McCormick 1966; Zaller 1988; cf. Finn 1972), although our demonstrations of branching effects in Studies 5, 7, and 8 are the first of their kind. Given the failure of a branching effect to appear in Study 6 and the failure of a labeling effect to appear in Study 5, we look forward to future studies assessing the validity of our conclusions. In the meantime, our results strongly suggest that verbal labeling and branching should be implemented whenever possible in surveys. Especially because fully labeled branching questions take *less* time to ask than partially labeled nonbranching questions, the former clearly seem preferable to the latter. Furthermore, researchers should be especially cautious when comparing the over-time consistencies or effects of attitudes if some of those attitudes were measured using branching and/or fully labeled formats and others were not.

Our findings suggest a reinterpretation of observed differences between party identification and policy attitudes in terms of over-time consistency in the NES surveys (Converse 1964; Converse and Markus 1979; Markus 1982; Sears 1983). The fully labeled branching method used to measure party identification in the NESs apparently enhanced its reliability relative to policy attitudes. Some of the observed differences in over-time consistency should therefore be attributed to differences in the reliability of the measures, not to stability of the underlying attitudes. Consequently, previous studies have probably overestimated the difference between the stabilities of party identification and policy attitudes.

Reliability affects not only the over-time consistency of attitude reports but also the amount of impact an attitude appears to have on other psychological variables. Therefore, although party identification has been shown in many analyses of NES data to be "the single most important determinant of individual voting decisions" (Kinder and Sears 1985, 686), its apparently stronger impact may be attributable at least partly to the fact that it has been measured more reliably. Thus, numerous previous studies may have underestimated the impact of policy attitudes on vote choice relative to the impact of party identification. The U.S. electorate

may therefore resemble political theorists' ideal responsible citizenry more closely than prior research has suggested.

Before accepting this conclusion, it is critical to assess the impact of a labeling and branching manipulation on full-rank regression equations predicting candidate preferences using a wide array of determinants (e.g., party identification, policy stands, perceptions of candidates' personalities, retrospective assessments of the nation's economic and international well-being, and so on). No study we know of has yet collected the necessary measures to permit such a comparison, and we look forward to having the opportunity to test this prediction directly when such data are available.

There is a striking parallel between our findings and those of Bishop, Tuchfarber, and Oldendick (1978; Bishop, Oldendick, and Tuchfarber 1978), Sullivan, Piereson, and Marcus (1978), and Brunk (1978). These authors explored the validity of an important argument made in *The Changing American Voter* (Nie, Verba, and Petrocik 1979): that Americans' views on public policy issues became dramatically more constrained as the result of the 1964 presidential election campaign. As is now widely recognized, Bishop, Tuchfarber, and Oldendick (1978; Bishop, Oldendick, and Tuchfarber 1978), Sullivan, Piereson, and Marcus (1978), and Brunk (1978), all demonstrated that Nie, Verba, and Petrocik's (1979) conclusion was almost certainly incorrect. Through controlled experiments that varied the wording and format of survey questions, these investigators demonstrated that the apparent increase in attitude constraint was almost certainly completely due to a change in the wording of the NES survey questions instituted in 1964.

The findings reported here regarding the impact of labeling and branching on item reliability have a similar quality. Like prior investigators, we have shown that a widely accepted truism is actually partly the result of a methodological artifact. Thus, we have provided the first direct evidence to support Achen's (1975) claim that at least some of the apparent unusual lability of policy preferences can be attributed to the survey questions used to measure them, rather than to the preferences themselves. And we have shown that investigators should pay close attention to survey question formats when testing substantive hypotheses. To ignore measurement procedures is to risk reaching inappropriate substantive conclusions.

Manuscript submitted 23 June 1992
Final manuscript received 3 December 1992

REFERENCES

Abramson, Paul R. 1983. *Political Attitudes in America: Formation and Change*. San Francisco: Freeman.

Achen, Christopher. 1975. "Mass Political Attitudes and the Survey Response." *American Political Science Review* 69:1218–31.

Aldrich, John H., Richard G. Niemi, George Rabinowitz, and David W. Rohde. 1982. "The Measurement of Public Opinion about Public Policy: A Report on Some New Issue Question Formats." *American Journal of Political Science* 26:391–414.

Alwin, Duane F. 1973. "Making Inferences from Attitude-Behavior Correlations." *Sociometry* 36:253–78.

Armstrong, J. Scott, William B. Denniston, and Matt M. Gordon. 1975. "The Use of the Decomposition Principle in Making Judgments." *Organizational Behavior and Human Performance* 14:257–63.

Bishop, George F., Robert W. Oldendick, and Alfred J. Tuchfarber. 1978. "Effects of Question Wording and Format on Political Attitude Consistency." *Public Opinion Quarterly* 42:81–92.

Bishop, George F., Alfred J. Tuchfarber, and Robert W. Oldendick. 1978. "Change in the Structure of American Political Attitudes: The Nagging Question of Question Wording." *American Journal of Political Science* 22:250–69.

Brunk, G. G. 1978. "The 1964 Attitude Consistency Leap Reconsidered." *Political Methodology* 5:347–59.

Campbell, Angus, Philip E. Converse, Warren E. Miller, and Donald E. Stokes. 1960. *The American Voter*. New York: Wiley.

Campbell, Angus, Gerald Gurin, and Warren E. Miller. 1954. *The Voter Decides*. Evanston, IL: Row, Peterson.

Converse, Philip E. 1964. "The Nature of Belief Systems in Mass Publics." In *Ideology and Discontent*, ed. David E. Apter. Glencoe, IL: Free Press.

———. 1966. "The Concept of a Normal Vote." In *Elections and the Political Order*, ed. Angus Campbell, Philip E. Converse, Warren E. Miller, and Donald E. Stokes. New York: Wiley.

Converse, Philip E., and Gregory B. Markus. 1979. "Plus ça Change . . . : The New CPS Election Study Panel." *American Political Science Review* 73:32–49.

Dahl, Robert A. 1956. *A Preface to Democratic Theory*. Chicago: University of Chicago Press.

———. 1989. *Democracy and Its Critics*. New Haven: Yale University Press.

Duncan, Otis D. 1975. *Introduction to Structural Equation Models*. New York: Academic Press.

Feldman, Stanley. 1989. "Measuring Issue Preferences: The Problem of Response Instability." *Political Analysis* 1:25–60.

Finn, R. H. 1972. "Effects of Some Variations in Rating Scale Characteristics on the Means and Reliabilities of Ratings." *Educational and Psychological Measurement* 32:255–65.

Fiorina, Morris P. 1981. *Retrospective Voting in American National Elections*. New Haven: Yale University Press.

Franklin, Charles H. 1983. "The Dynamics of Party Identification." *American Political Science Review* 77:957–73.

Jennings, M. Kent, and Gregory B. Markus. 1984. "Partisan Orientations over the Long Haul: Results from the Three-Wave Political Socialization Panel Study." *American Political Science Review* 78:1000–18.

Jennings, M. Kent, and Richard G. Niemi. 1981. *Generations and Politics*. Princeton: Princeton University Press.

Kinder, Donald R., and David O. Sears. 1985. "Public Opinion and Political Action." In *Handbook of Social Psychology*, ed. Gardner Lindzey and Elliot Aronson. New York: Random House.

Krosnick, Jon A. 1988. "The Role of Attitude Importance in Social Evaluation: A Study of Policy Preferences, Presidential Candidate Evaluations, and Voting Behavior." *Journal of Personality and Social Psychology* 55:196–210.

———. 1991a. "Response Strategies for Coping with the Cognitive Demands of Survey Questions." *Applied Cognitive Psychology* 5:213–36.

———. 1991b. "The Stability of Political Preferences: Comparisons of Symbolic and Nonsymbolic Attitudes." *American Journal of Political Science* 35:547–76.

Madden, J. M. 1960. "A Comparison of Three Methods of Rating Scale Construction." USAF WADD Personnel Laboratory Technical Note 60-262.

Markus, Gregory B. 1982. "Political Attitudes during an Election Year: A Report on the 1980 NES Panel Study." *American Political Science Review* 76:538–60.

Nie, Norman H., Sidney Verba, and John R. Petrocik. 1979. *The Changing American Voter*. Cambridge: Harvard University Press.

Pennock, J. Roland. 1979. *Democratic Political Theory*. Princeton: Princeton University Press.

Peters, David L., and Ernest J. McCormick. 1966. "Comparative Reliability of Numerically Anchored versus Job-Task Anchored Rating Scales." *Journal of Applied Psychology* 50:92–96.

Rosenthal, Robert. 1988. *Meta-Analytic Procedures for Social Research*. Beverly Hills: Sage.

Sears, David O. 1983. "The Persistence of Early Political Predispositions: The Roles of Attitude Object and Life Stage." In *Review of Personality and Social Psychology*, vol. 4, ed. Ladd Wheeler. Beverly Hills: Sage.

Sullivan, John L., James E. Piereson, and George E. Marcus. 1978. "Ideological Constraint in the Mass Public: A Methodological Critique and Some New Findings." *American Journal of Political Science* 22:233–49.

Wattenberg, Martin P. 1984. *The Decline of American Political Parties, 1952–1980*. Cambridge: Harvard University Press.

Zaller, John. 1988. "Vague Minds vs. Vague Questions: An Experimental Attempt to Reduce Measurement Error." Presented at the annual meeting of the American Political Science Association, Washington, DC.