

NORBERT SCHWARZ +
SEGMAR SCHUMAN eds

Context Effects in
SOCIAL + PSYCHOLOGICAL
RESEARCH

Springer-Verlag

1991

NY

12 Thoughts on the Nature of Context Effects

Tom W. Smith

Despite over 40 years of study, question order is probably the least developed and most problematic aspect of survey research. As Schuman and Presser (1981) remarked in their work on survey methodology:

Overall, order effects . . . constitute one of the most important areas for methodological research. They can be very large [and] are difficult to predict At this point research needs to be aimed not merely at producing more examples, but at understanding why those already obtained occur. (p. 77)

This perplexity is shared by Bradburn (1983), who observes, "No topic in questionnaire construction is more vexing or resistant to easy generalization than that of question order" (p. 302), and by Groves (1989), who notes that "there seems to be no general theory that predicts when such effects are to be expected and when they should not be expected" (p. 479). There is a temptation to blame our collective befuddlement on a dearth of experimental studies. Although we, like *Oliver Twist*, would like "more," the paucity of data is not the main cause for our ignorance. There have been nearly 100 studies of order effects, most involving split-ballot experiments.

First and most fundamentally, understanding has been limited because the topic is extremely complex. It now appears that there are many distinct types of order effects. Until recently we have been like 19th century physicians, who used the term "a cancer" to cover many separate diseases. We are now only beginning to distinguish, sort out, and study the different types of order effects and their causes. We are not even sure at this point if we have identified the main causal processes or correctly specified the major types of order effects. We are beginning to realize that a knowledge of social psychology (e.g., attitude change) and cognitive psychology (e.g., memory recall and linkage) will be required to understand order effects.

Second, development has been hampered by an atheoretical focus. Most early studies have lacked explicit (and a number even implicit) explanations for the effects under investigation. The development and testing of competing hypoth-

eses has typically been ignored, and even when the previous literature is cited by later studies, there is often no cumulateness of research. Studies are cited as examples, but we have not tended to learn from these examples.

Third, there has been a major underanalysis of existing empirical data. The majority of experiments merely compare the marginal distribution of B under orders AB and AB. Reciprocal marginal effects, interitem associations, conditional effects, and interactions with other variables have rarely been examined.

In brief, we have been trying to understand a complex problem without adequately applying either the theoretical or empirical tools of the social scientific method. As a result, we have been able to demonstrate repeatedly the existence or nonexistence of various particular order effects with little cumulative understanding of the causes and conditions involved.

This chapter (a) examines the use of conditional order effects as a method for understanding the nature of context effects, (b) considers how common context effects are, (c) evaluates the related issues of scattering and buffering, and (d) reviews various systems for classifying context.

Conditional Order Effects

Conditional order effects are one of the most commonly overlooked yet most important aspects of context effects. Almost all studies prior to Schuman and Presser's (1981), as well as many since, have assumed that it is the prior question or questions themselves that have induced order effects in subsequent questions. This holistic assumption appears likely for certain types of order effects (see section later in this chapter on Classifications and Causes) but not for other types. Implicitly (and rarely explicitly) in the early literature, there is an indication that the order effect rests not only on the context of a prior question but also on how one responded to the antecedent question. This interaction between question order and response to the antecedent question is what I call a "conditional order effect." I focus on this aspect of order effects because (a) I believe that conditional effects are common among order effects and (b) understanding the conditional relationship between antecedent and subsequent responses greatly facilitates comprehending the nature and causes of context effects.

Prior to the work of Schuman and Presser (1981), not one study tested for conditional effects. This makes conditional effects the most neglected aspect of order effects. In contrast, other aspects of order effects beyond unidirectional, marginal effects, such as interitem associations, reciprocal marginal effects, and interactions with other variables, have been measured in various studies. To study conditional order effects, I was able to draw on three examples from Schuman and Presser: general and specific abortion, Communist and American reporters, and general and specific job discrimination; five examples from the General Social Surveys (GSSs): tax and spending, alienation and institutional confidence, marital and general happiness, national service for men and women, and anomia; one example from the Greater Cincinnati Surveys: political interest

and congressional knowledge; and four examples from NORC's Chicago context effects surveys: welfare spending and economic individualism, welfare spending and government responsibility, Cuba and aid to the contras, and Vietnam and aid to the contras. The GSSs are based on full probability personal interviews and the rest on random-digit-dialed (RDD) telephone surveys. The Schuman and Presser and GSS experiments are based on samples of the national adult population conducted between 1976 and 1982 by, respectively, the Survey Research Center at the University of Michigan or the National Opinion Research Center at the University of Chicago (for more details, see Schuman & Presser, 1981, and Davis & Smith, 1989). The Greater Cincinnati experiments were conducted in the Cincinnati metropolitan area in 1983-84 (Bishop, 1987), and the NORC Chicago context effects surveys were carried out in Chicago in 1987 (Tourangeau, Rasinski, Bradburn, & D'Andrade, 1989a).

Table 12.1 shows six cases in which context effects were conditional on responses to the antecedent question (see Davis & Smith, 1989, and Schuman & Presser, 1981, for wordings). In the first example, the overall context effect is for the appearance of the marital happiness question immediately before the general happiness question to increase general happiness.¹ Looking at the conditional context effects, we see that the effect is largely confined to those rating their marriages as very happy. Mentions of marital happiness increase general happiness, since most married people rate their marriage as very happy; but among the unhappily married, there is no nuptial bliss to spread to general happiness.

In the second example, placing alienation items before confidence items reduces the confidence rating of major companies. This effect is, however, entirely confined among those who agreed with the proposition that "the rich get richer and the poor get poorer." Similarly, asking about allowing Communist reporters to gather news in the United States first reduces support for allowing an American reporter to cover a Communist country such as Russia only among those opposed to allowing Communist coverage of the United States.² Likewise, in the tax/spending example, fiscal conservatives (people rejecting most current spending levels as too high) do not vary their opinions on taxes, whereas spending moderates and liberals are less likely to object to taxes after the spending items. Next, people who favor national service for women do not differ in their attitudes toward national service for men by whether the item on men or women is asked first, but those opposing national service for women show less support for national service for men when the men question follows the women question. Lastly, the asking of two anomia items in a row increases the anomie responses

¹A result that is at odds with Schuman and Presser's similar experiment (1981; Schuman, Presser, & Ludwig, 1981).

²The Communist/American reporters example is actually more complicated than the others because the marginal effects are reciprocal. As a result, the distribution of the conditional controls varies by order.

to the second item. This response occurs among those who agree with the first statement that the lot of the average man is getting worse and does not occur among those disagreeing with that statement. (For more analysis of this experiment, see T. W. Smith, 1983b.)

TABLE 12.1. Six Examples of Conditional Context Effects

Item	Order	Context Effect
General Happiness by Marital Happiness by Order		
	Marital/General	General/Marital (Order 1 - Order 2)
General happiness (% very)	Marital Happiness = Very Happy	
	56.1 (421)	47.5 (177) 8.6
Marital Happiness = Not Very Happy		
	11.5 (192)	8.8 (91) 2.7
Confidence in Major Companies by Alienation by Order		
	Alienation/Confidence	Confidence/Alienation (Order 1 - Order 2)
Major companies (% great deal)	Rich Get Richer = Yes	
	11.9 (528)	22.6 (541) -10.7
Rich Get Richer = No		
	38.9 (175)	39.2 (169) 0.7
American Reporters by Communist Reporters by Order		
	American/Communist	Communist/American (Order 1 - Order 2)
American reporters (% allow)	Communist Reporters = Allow	
	99.0 (100)	96.2 (130) 2.8
Communist Reporters = No		
	21.6 (74)	40.0 (40) -18.4
Tax Approval by Spending Preferences by Order		
	Spend/Tax	Tax/Spend (Order 1-Order 2)
Tax (% taxes too high)	Spend Scale = Antispending	
	59.7 (144)	61.0 (141) -1.3
Spend Scale = Not Most Antispending		
	68.0 (400)	49.4 (389) 18.6

(Table 12.1 continued)

TABLE 12.1. Continued

Item	Order	Context Effect
National Service for Women by National Service for Men by Order		
	Men/Women	Women/Men (Order 1 - Order 2)
(% favor, strongly favor for men)	National Service for Women = Favor	
	98.3 (464)	98.9 (443) 0.6
National Service for Men = Oppose		
	39.0 (246)	28.4 (282) -10.6
Anomia by Anomia		
Clustered (Lot Getting Worse/ Not Fair to Have Child)		
	Scattered	(Order 1 - Order 2)
Not fair to have child (% agree)	Lot Getting Worse = Agree	
	50.9 (475)	31.6 (455) -19.3
Lot Getting Worse = Disagree		
	21.8 (211)	19.8 (232) -2.0

For two of these examples, I was able to examine conditional effects in greater detail by looking at seven levels on the alienation scale and four spending levels (Table 12.2). First, in both cases the overall order effect (less confidence in business after alienation items and less opposition to taxes after spending questions) is not merely absent under certain conditions but reverses at one pole. The outlook of the extreme antispending and unalienated groups differs so much from the majority that the spending and alienation items have an opposite impact on them than for the majority. This means that the gross order effect across individuals is substantially greater than the net effect observed among the aggregate population.

The second similarity is more surprising. The largest order effect in the main direction does not occur at the opposite pole on alienation but in the middle. The middle conditional order effect is also large on the tax/spending example. In both cases this effect occurs among the median group, those with 3 agrees and 3 disagrees on the alienation scale and those with an average score of 2 (spending about right) on the 11 spending items. I hypothesized that the effects might increase among the median groups because those groups contained a large share of people with weak attitudes on the issues whose median scores were more a product of nonattitudes and random responding than a reflection of a considered middle position. Being without fixed attitudes, they were more swayable by

TABLE 12.2. Detailed Conditional Effects

Item	Order		Context Effect (Order 1 - Order 2)
	Confidence in Major Companies by Alienation/ Confidence	Confidence/ Alienation	
Confidence in major companies (% great deal)	Alienation Scale = 0		
	51.5 (68)	40.0 (40)	11.5
	Alienation Scale = 1		
	27.9 (61)	34.9 (83)	-7.0
	Alienation Scale = 2		
	20.9 (86)	34.5 (94)	-13.4
	Alienation Scale = 3		
	10.1 (89)	33.0 (94)	-22.9
	Alienation Scale = 4		
	24.1 (87)	28.7 (108)	-4.6
Alienation Scale = 5			
10.0 (110)	13.5 (104)	-3.5	
Alienation Scale = 6			
6.2 (97)	12.1 (91)	-5.9	
Tax (% taxes too high)	Tax Approval by Spending Preferences by Order		
	Spend/Tax	Tax/Spend	(Order 1 - Order 2)
	Spend Scale = Most Antispending		
	57.8 (90)	65.2 (89)	-7.4
	Spend Scale = Low Spending		
	69.8 (182)	49.4 (168)	20.4
Spend Scale = Moderate Spending			
65.4 (208)	55.0 (191)	10.4	
Spend Scale = High Spending			
65.6 (64)	40.2 (82)	25.2	

question order (Tourangeau & Rasinski, 1988; Tourangeau, Rasinski, Bradburn, & D'Andrade, 1988).

I tried to check this by examining whether this group showed less interest, knowledge, or involvement. The median group did not overrepresent less educated respondents or those giving "don't knows" to other attitude questions. On the spending questions, however, the median group had the highest level of non-voting (32.8% vs. 22.3% for everyone), but on the alienation items, no difference appeared. These minimal results probably occurred because the median group contained both random responders and those with moderate positions and because of the difficulty of finding general items that would predict random responding to a particular scale. The one confirmation on the voting item may indicate that my explanation for why middle order effects were high is plausible.

The next two examples (Tables 12.3 and 12.4) show conditional effects occurring in combination with general context effects. In Table 12.3, the general

TABLE 12.3. Respondents Who Follow Politics "Most of the Time," by Congressional Knowledge and Context (%)

Congressional Knowledge	Order		Context Effect
	Follows/ Cong. Items	Cong. Items/ Follows	
Knows about Congress = Both (2)	74.0 (50)	66.7 (36)	-7.3
Knows about Congress = Partial (1)	55.3 (199)	38.3 (162)	-16.9
Knows about Congress = Neither (0)	30.7 (651)	21.4 (669)	-9.3

Source: Adapted from Bishop (1987, Table 3). Order compares Forms A and B and collapses over and ignores item about knowing governor.

effect is for prior questions asking about the actions and votes of one's representative in Congress to lead to fewer reports of following politics and public affairs "most of the time." This effect occurs regardless of how much knowledge one professed. The context effect is not uniform, however, but also conditional, being much larger among those claiming partial knowledge than among those with full or no knowledge. This resembles the pattern appearing in Table 12.2, with the median group showing a larger effect than the extremes. Unlike the other cases, the trigger questions measure knowledge rather than an attitude. Perhaps those responding to the two difficult congressional questions include a number of labile respondents who had randomly responded to the knowledge questions (for more on this experiment, see Bishop, 1987).

The example in Table 12.4 shows the results from four of the NORC Chicago context effects studies. In each case, there are main effects (support for welfare spending is shifted by prior items on both economic individualism and government responsibility and favoring contra aid is influenced by the earlier items on Cuba and Vietnam) as well as conditional effects. The context effects are typically twice as large among those with high agreement with the trigger items as among those with low agreement with these items. For example, among those with low agreement with the Vietnam items, the context effect is -9.1 percentage points, whereas those with high agreement show an effect of -19.1 percentage points (for more analysis of these experiments, see Tourangeau & Rasinski, 1988; Tourangeau et al., 1988, 1989a).³

The final two examples (Table 12.5) show no evidence of conditional order effects, but actually both underscore the importance of checking for these specifications. As Schuman and Presser note, the lack of a conditional effect on the abortion questions is surprising, since their prime explanation of the effect (a

³These results are contrary to those reported by these authors. They generally report no such conditional effects. For example, "Similarly, the effects of context did not depend on the respondent's initial opinion about the target issue" (Tourangeau et al., 1988, p. 30).

TABLE 12.4. Endorsement of Welfare and Nicaragua Target Items, by Context and Level of Agreement with the Context Items (%)

Context Set	Low	High
Favor Increased Welfare Spending		
Agreement with Government Responsibility Items		
Economic individualism	42.4	52.5
Government responsibility	52.3	78.1
Context effect (bottom - top)	+9.9	+25.6
Agreement with Economic Individualism Items		
Economic individualism	56.5	40.7
Government responsibility	66.8	58.8
Context effect (bottom - top)	+10.3	+18.1
Favor Increased Welfare Spending		
Agreement with Cuba Issue		
Cuba	26.2	57.9
Vietnam	17.2	38.5
Context effect (bottom - top)	-9.0	-19.4
Agreement with Vietnam		
Cuba	40.4	41.6
Vietnam	31.3	22.5
Context effect (bottom - top)	-9.1	-19.1

Source: This table is adapted by permission from Tourangeau et al. (1989a, Table 4). Copyright © by Academic Press. The *N* for each row is approximately 500.

Note: Respondents received both sets of context items (e.g., the items on economic individualism and those on government responsibility), with one set coming before the relevant target item and the other set coming afterward. The context set variable indicates the items that preceded the target.

subtraction effect) implies such an effect. They argue that people who are presented with the popular, specific reason for abortions in case of possible defects in the unborn child first tend to exclude this reason from the subsequent general abortion question and thereby lower their support for the general abortion item. This scenario works nicely for the majority of people who approve of abortions in cases of possible birth defects, but it fails to explain why people who opposed abortion for birth defects are also less likely to approve of general abortions when the specific birth defect item comes first. Presumably since birth defects have been rejected as a good reason for an abortion, there is no positive component to subtract out of the general abortion question. Either there is a general

TABLE 12.5. Six Examples of Conditional Context Effects

Item	Order	Context Effect
General Abortion by Specific Abortion by Order		
	Specific/General	General/Specific (Order 1 - Order 2)
General abortion (no more children = Yes)	Specific Abortion (Defect) = Yes	
	56.1 (246)	69.2 (253)
	Specific Abortion (Defect) = No	
	6.4 (47)	19.2 (52)
General Job Discrimination by Specific Job Discrimination by Order		
	Specific/General	General/Specific (Order 1 - Order 2)
General discrimination (in principle) = Favor	Specific Discrimination (Avoid Friction) = Favor	
	18.7 (32)	13.3 (30)
	Specific Discrimination (Avoid Friction) = Oppose	
	9.6 (157)	3.2 (158)

explanation other than the subtraction effect proposed by Schuman and Presser, or we have two distinct conditional effects that happen to be equal in magnitude.

Alternative explanations include a contrast effect. The general reason may not seem as attractive when compared to the highly attractive birth defect reason, and therefore fewer people may endorse the general abortion question. This contrast effect could work either among people opposed to abortions for birth defects or among those in favor of it, since even those opposed to abortions for birth defects might recognize it as a better reason than general abortion and therefore reduce their approval of the general item.

Another possible explanation has similarities to the subtraction effect, a redefinition effect. When the general question appears first, some people think of the various reasons for not having another child, and since some of the reasons are attractive (e.g., the prevention of birth defects), they approve of the general abortion question. When it comes second, they realize that it does not mention birth defects and may infer that it does not include any other extenuating circumstances either. It thus changes from being a general abortion question to being a specific question about unwanted children. The specific/general ordering clarifies that the so-called general question does not include any extraordinary reasons for not wanting another child but simply a desire to avoid more children. Thus, even someone opposed to abortion for birth defects would be less likely to support general abortion not because birth defects are excluded from the question but because the question is seen as excluding all special circumstances. Since the context redefines what the general question is asking about, it changes how

everyone responds to the question regardless of their attitude on the birth defect item.

Although either of these general explanations may explain the lack of an interaction within the birth defect question, it is also possible to come up with particular explanations for those opposed to abortions for birth defects. From a Guttman scaling perspective, those who say "no" to abortion for birth defects but "yes" to abortion for preventing more children represent an error group.⁴ Perhaps these cases do represent error by people who are confused by or inattentive to the abortion question. Although the specific-to-general (easy-to-hard) order reduces error, the opposite order permits more random error on the general question. Perhaps the appearance of the general question second allowed respondents more time to sort out their thoughts on abortion and therefore to give consistent rather than inconsistent response patterns. This would leave among the error cases those most confused about the abortion issue and a group whose true pattern deviated from the predominant pattern (e.g., those who thought defective children were God's special children and a blessing in disguise but that unwanted normal children would be raised without love and thus best prevented).

The situation about job discrimination is similar to abortion. No conditional effect is observed, but as Schuman and Presser note, this is counter to the consistency explanation suggested by the marginal shifts. I shall not go through possible alternative explanations for the absence of a conditional order effect but instead reiterate that the absence of such an effect is often as informative as its presence.

In 8 of 10 examples available, order effects were concentrated in whole (6 cases) or in part (2 cases) among certain categories of the antecedent question. It was not the mere mention of a prior topic that induced a marginal shift in the subsequent question but a respondent's position on the antecedent variable and the order that induced the order effect. In fact, from the tax/spending and alienation/confidence items, we see that even the direction of the order effect is dependent on the position on the antecedent item. This information can not only be used for a better understanding of the particular observed order effects (along with other empirical analysis of reciprocal marginals, interitem correlations, and interactions with other variables) but also perhaps allow a refined classification of order effects, better theory, and improved predictions of when order effects are likely.

Of course, context effects that are conditional on one's attitudes toward prior topics but not conditional on one's responses to prior questions may also occur. As I shall suggest in a later example, a prior question may lead one to access memories and beliefs that specify later questions (see section on Classifications

⁴Looking at the six abortion items on the GSS, which include the two items used by Schuman and Presser, we find that the general abortion item is the hardest item to approve, while the birth defect item is the second easiest. The coefficients of reproducibility and scalability are .94 and .81.

and Causes), but the prior question may not be framed in a way that allows recording expressions of those memories and beliefs so as to permit the measurement of a conditional order effect. It is unknown whether actually expressing an attitude in response to a prior question or merely accessing (but not expressing in a response to a prior question) the relevant memories and beliefs would create a similar context effect. One suspects, however, that actual expression of the conditional attitude in a prior question might exert a greater effect than activation without explicit expression.⁵

Commonness of Context Effects

There is some disagreement on how common context effects are. Tourangeau et al. (1988) conclude that "the literature on survey context effects may create the impression that such effects are relatively rare, involving items on a few scattered issues. These results here indicate otherwise" (pp. 22-23). This impression of pervasiveness is supported by numerous instances in which changes in question order have upset time series or caused other undesired measurement variations (Astin, Green, Korn, Schalit, & Berz, 1988; Cowan, Murphy, & Wiener, 1978; Gibson, Shapiro, Murphy, & Stanko, 1978; T. W. Smith, 1986b, 1988c; Turner & Martin, 1984). Schuman and Presser (1981), on the other hand, reach a conclusion that at least differs in emphasis: "Question-order effects are evidently not pervasive, . . . but there are enough instances to show they are not rare either" (p. 74). This nonpervasive impression is supported by numerous failures to produce context effects in experiments designed to do so (Schuman & Presser, 1981; T. W. Smith, 1983a; Turner and Martin, 1984) and by the ability of different houses to produce similar marginals when the same questions, but different question content (and other variations), existed (Turner & Martin, 1984; T. W. Smith, 1978, 1982).

Two studies have conducted general searches for context effects.⁶ Schuman and Presser (1981) examined the 1971 Detroit Area Study (DAS). The DAS used split ballots in order to accommodate various experiments in either question order or wording. They looked at 113 attitude items that were not the designed objects of these experiments but appeared after the experiments and thus varied in

⁵In terms of the cognitive framework of Tourangeau and Rasinski that we explore later, there are several ways that conditional order effects could be created. Even if no relevant attitude was directly expressed, a retrieval carryover effect could occur if a prior question triggered selective memory sampling and people differed in their affect toward the primed memories. In addition, the expression of a relevant prior attitude could create a conditional order effect by causing judgmental carryover or consistency editing during the response selection stage.

⁶Also see Bradburn and Mason (1964), which tested for 14 differences in marginals across four forms and found no statistically significant variation.

context due to the prior experiments. Apparently using simple random sample (SRS) assumptions, they found eight significant differences at the .05 level, just two above what chance would predict. Their inspection of these eight suggested that three probably represented real effects and the rest were due to sample variation.

I have examined the 1988 GSS (T. W. Smith, 1988a). In 1988 the GSS switched from an across-years rotation scheme to a within-year split-ballot design (T. W. Smith, 1988b). That meant that three split ballots were employed. Each ballot represented a year under the old across-years rotation scheme. Demographics typically appeared on all three ballots, and attitudes and other items usually appeared on two of the three ballots. In the vast majority of cases, the items appearing on different ballots appeared in very different orders. I tested both for context effects across the ballots and for evidence of context effects by grouping together earlier years that largely duplicated the same orders that appeared in the 1988 ballots. Among 358 questions that varied in context across the ballots, 9.2% were found to vary significantly using SRS assumptions, but only 3.6% varied significantly when adjusted for design effects. Close examination of the 14 statistically significant adjusted results suggests that 6 probably represent real context effects and the remaining 8 are chance occurrences.

These studies suggest that unanticipated context effects might occur once out of every 40–60 questions. However, this is probably an underestimate, since on the GSS, and presumably on the DAS, batteries of questions on one topic (e.g., the seven abortion items) were asked in a block and not varied across ballots. Since context effects are most likely to occur between closely related items, the failure to vary items experimentally within topical blocks probably underestimates the frequency of context effects.⁷

Buffers, Scattering, and Context Effects

Almost all early research on context effects placed the experimental variation in order immediately prior to the target item. In recent years, however, several studies have been conducted that have varied the placement of the trigger question(s). One approach inserts a buffer of unrelated items between the trigger and the target, and the other presents the trigger items either in a block or scattered among unrelated items. In the buffering approach, investigators have taken a well-known context effect and tested its power by inserting a buffer of items between

⁷The abortion example discussed above is a prime example of what can occur when the order within such a block is disturbed. See also Schuman and Presser's (1981) discussion of same. For other examples, see Astin et al. (1988) and T. W. Smith (1984). I suspect that within-scale context effects are rather common, probably even typical. Since such scales tend to be replicated as units without changes to their internal order, these effects are rarely studied, and whatever context effects exist generally remain fixed across administrations of the scale.

the trigger and the target (Bishop, 1987; Schuman, Kalton, & Ludwig, 1983). These studies indicate that context effects are quite robust and work with little or no diminution even when a large buffer intervenes. In the block/scattered approach, the investigators test whether concentration in a block is needed to affect respondents or perhaps is so obvious that it creates a backfire effect (Tourangeau et al., 1988, 1989a). Despite some early indications that scattering might be more effective than blocking, Tourangeau and his colleagues now believe that scattering diminishes and may eliminate context effects. Their meta-analysis of the buffer and scattered/blocked experiments showed that intervening items do diminish context effects (Tourangeau et al., 1988).

This conclusion is supported by my research on context effects that suggests a first-only effect (T. W. Smith, 1981a, 1988a). In this research, I found three examples of context effects influencing only the first item on subsequent scales. For example, alienation items reduced the confidence rating of the first item in the 13-item confidence items but had no significant impact on the distributions of the following 12 items (T. W. Smith, 1981a). This suggests not only that scattering or a wide buffer can reduce a context effect but that a single item may absorb the effect. The very robust effects detected by some investigators may be because they tested particularly strong context effects, and these results may be the exceptions rather than the rule.

Classifications and Causes

In order to advance in the study of context effects, we need to develop theories about what causes such effects. Two approaches have been used that might facilitate this process. The first is the classification of context effects into different types according to their cause and effect. The second is to delimit the cognitive steps involved in answering questions and to relate how context might operate during each of these steps. Although these two approaches have developed independently and are distinct, they overlap. The classification approach largely evolved out of an attempt to explain existing or known context effects by applying relevant social-psychological and cognitive theories. The cognitive-steps approach came from a general attempt to apply cognitive theory first to the survey process in general and then to context effects in particular.

Schuman and Presser (1981) and Bradburn (1983; Bradburn & Mason, 1964) have formulated two similar classifications schemes for question-order effects⁸ (see Table 12.6). Both refer to psychological or cognitive processes by which order influences subsequent questions. Schuman and Presser's classification is more detailed and more hierarchically organized than Bradburn's but mainly differs by using question-type distinctions (part and whole) within the consistency

⁸Here, as elsewhere in this chapter, we exclude the related matter of response-order effects (see Schuman & Presser, 1981, pp. 56–74).

TABLE 12.6. Classification of Order Effects

Schuman and Presser	Bradbun
I. Context effects (transfers of meaning)	
A. Part-part consistency	
1. Normative principles	
2. Logical inference	1. Consistency
B. Part-whole consistency	
C. Part-part contrast	
D. Part-whole contrast	
1. Subtraction	2. Redundancy
2. Simple contrast	
E. Saliency	3. Saliency
II. Sequence effects (more mechanical types of artifacts)	
A. Rapport	4. Rapport
B. Fatigue	5. Fatigue
C. Initial frame of reference	

Source: Schuman and Presser (1981), Bradburn and Mason (1964), and Bradburn (1985).

and contrast categories and the addition of the initial frame of reference and simple contrast classes.

Tourangeau, Rasinski, and others (Tourangeau & Rasinski, 1988; Tourangeau et al., 1988; Strack & Martin, 1987) have more recently described the various steps and processes involved in answering survey questions and how context could effect each step and process. Tourangeau and Rasinski (1988) describe the four steps in answering a survey question: (a) interpretation, (b) retrieval, (c) judgment, and (d) response selection. At each stage they posit two types of context effects: carryover and backfire.⁹ Carryover effects involve the usually automatic or unconscious influence of prior questions on subsequent questions (except for editing during the response selection stage). In some general sense, a prior question shapes responses to a later question during one or more of the answering stages. Backfire effects are a more conscious rejection of the influence of prior questions when answering later questions. In addition to these eight types of possible context effects, Tourangeau and Rasinski subdivide the response selection stage into the processes of mapping attitudes into response categories and editing for the sake of either consistency or self-presentation.

The Tourangeau-Rasinski approach has the decided advantage of grounding context effects in cognitive theories. Their work is an excellent example of intellectual, hybrid vigor, since it uses the results from psychological experiments to illuminate survey research. But in explaining and illustrating their differing types of context effects, they may draw too heavily from diverse literatures in experimental psychology at the expense of underexamining survey research's own experiments on context effects.

⁹The authors sometimes refer to these as assimilation and contrast, respectively.

In the following sections, I shall use a slightly modified version of the Schuman-Presser/Bradburn classifications¹⁰ to review the extant survey literature on context effects (see T. W. Smith, 1986a, for the literature) and then consider how this standard scheme meshes with the cognitive approach of Tourangeau-Rasinski.

Order effects come in many variations. Some context effects are unrelated to the substance of the prior questions, others are related to the prior substance but not to prior responses, and still others are related to both the substance and one's response to the prior question. First, there are Schuman and Presser's sequence effects (also called "position effects"). These are sometimes described as "mechanical" and are believed to be completely unrelated to the substance of the preceding question(s). A rapport effect argues that a more trusting and open exchange of information occurs after the interview has developed. Less mentioned and perhaps sometimes subsumed under rapport effects are learning effects. Learning effects suggest that respondents, in general, learn their role as respondents better as the interview unfolds and in particular become more familiar with response scales and other tasks (e.g., the use of 7-point scales or sorting tasks). This reduces measurement error. At the opposite pole, a fatigue effect stipulates that, after a long series of questions, a respondent grows tired and gives less complete and more perfunctory answers. Another less commonly mentioned position effect is what Schuman and Presser (1981, pp. 51-52) call an "initial frame of reference effect." Within a battery of questions rating or comparing topics on a common criterion, an item will tend to receive either its lowest or highest mean rating when it appears first.

Second, there are what Schuman and Presser call "context effects," which involve some transference of meaning between the antecedent question and the subsequent question. Some of these context effects depend only on the topics raised in the prior questions and not on a respondent's affect toward or responses to these items. One example is a stimulation effect (akin to priming) in which questions about a subject stimulate more reports of behavior related to or interest in the topic. For example, attitude questions about crime lead to more reports of criminal victimization (Cowan et al., 1978; Gibson et al., 1978), and questions about politics increase reported levels of interest in politics.¹¹ Three quite distinct explanations have been offered for these increases: improved memory search leading to more complete reports, increased telescoping of behaviors causing exaggerated reports, and intentional exaggeration because of role fulfillment pressures. Although improved memory search is usually the favored explanation, it is quite possible that all three processes can be at work either in different situations or even simultaneously in the same situation (e.g., some of the increased

¹⁰The basic categories are employed, but (a) some additional refinements and distinctions are added, along with some new terminology, and (b) the part-whole distinction is not utilized (T. W. Smith, 1986a).

¹¹But for an exception, see Bishop, Oldendick, & Tuchfarber (1982, 1984a).

crime reports may come from a more thorough memory dragnet, whereas some come from increased telescoping).

Another effect that depends on the substance of prior questions is a redefinition or clarification effect (similar to redundancy and subtraction effects). For example, as part of a series of questions about the brand of washer, TV, etc., you own, an inquiry about "And what kind of car do you own?" would elicit more model names than the same question appearing alone, which would get more references to vans, sedans, convertibles, etc. Similarly, I posited above that the general-specific abortion effect might involve a redefinition of the general question. When the redefinition effect eliminates a specific element from the subsequent question, we have a subtraction effect as discussed earlier.

Closely related to the redefinition effect is redundancy. As Bradburn describes it, a person mentioning certain behaviors at an earlier point may consider it repetitive to mention them again. The respondent may believe that these elements are excluded from the subsequent question (redefinition) or simply be reluctant to go over the same ground twice even if the respondent realizes that the same information is applicable to the later questions.

Finally, simple contrast effects may fall into this category. Here one judges the desirability of the second question in light of the first. If the first represents a highly positive situation and the other a less attractive situation, the relative merit of the second item may seem even less because it is contrasted to the first and pales in comparison. This effect necessitates that a respondent recognizes a contrast between the desirability of two propositions but not necessarily that he endorses the attractive proposition. One need only recognize that, in general, such a distinction is seen.

Next there are context effects that depend not only on the substance of the prior question but also on responses to the antecedent question constraining response to the subsequent question. One such constraint or consistency effect involves the establishment of a normative principle between two questions. This is exemplified by the Communist/American reporters question. This type probably represents the strongest of context effects and usually, if not always, will cause reciprocal marginal effects (i.e., both A and B distributions will differ in orders AB and BA).

A second constraint effect establishes a logical connection between questions. This would include the tax/spending example. Although not too distant from normative effects (especially if we consider logic as a norm) and also involving, like the former, a conscious attempt to bring responses into line, the logical connection effect does not rest on a general social norm separate from the main substance of the items.¹²

Next comes a rather large and fairly amorphous category of focus effects (similar to salience effects). These focus attention on some topic that relates to the subsequent question. Questions about children preceding an abortion question

might reduce support for abortion, since the salient images of children might focus attention on unborn children rather than on women when considering the abortion questions.¹³ Unlike logical connection effects, focus effects do not come from strictly logical propositions but rather from more subtle pressures and inclinations, and the impact is seen as working through memory access rather than through conscious reconciliation of response patterns.

It is, however, often difficult to determine whether responses involve conscious logical constraint ("I am very happily married. My marriage is the most important part of my life. Therefore, my life is very happy") or patterns of cognition (in thinking about general happiness, R has most ready access to the marital happiness memories that have just been recalled). In either case, being very happy on marriage will lead to increased reports of happiness on the general question, but the causes or processes are not the same. In the former case, general happiness responses are being consciously reconciled with the prior marital happiness response, which comes from the accessed memories of marital happiness, whereas in the latter case, the effect comes directly from the memories.

Third, although involving conditional effects in a general sense, it may not be possible to demonstrate conditionality because the antecedent questions may not have an item that explicitly records the attitude that specifies the order effect. A focus effect is conditional in that it is what you feel toward the topic covered by the antecedent question that determines your subsequent response. This may not be discernable, since the antecedent question may not inquire about feelings toward the topic. For example, in the classic dress-advertising example (American Marketing Association, 1937), "questions regarding dresses" preceded attitudes toward advertising. I do not know just what dress questions were asked but suppose that these questions covered such matters as place of purchase, styles favored, and the like. Subsequent attitude questions revealed that after the dress questions (a) ratings of advertising was more favorable and (b) dress advertising was the main type of advertising thought of. The factor that leads the increased focusing on dresses to improve advertising ratings is that women like dresses and as a result presumably like dress advertisements. Among the presumably small proportion of women who disliked clothes in general or dresses in particular, we would presumably not find an increase in favorable ratings of advertisements. Although there are distinct differences in the processes involved in these two classes of effects, they are differences of degree and specific examples may involve blends of both.

Table 12.7 compares these standard categories of context effects on the Tourangeau-Rasinski scheme. The fitting of the standard types into this framework was a difficult, but useful, exercise. It revealed strengths and limitations of both systems, showed when conceptualizations were similar and when they were divergent, and raised the possibility of developing a better overall understanding of order effects by drawing on elements of both classifications.

¹²For example, where context failed to induce logical constraint, see T. W. Smith (1981a, 1981c).

¹³I found little support for this particular example (T. W. Smith, 1983a).

TABLE 12.7. Comparison of the Tourangeau and Rasinski Categories with Traditional Classifications of Context Effects

Question-Answering Steps	Reactions to Prior Questions	
	Carryover	Backfire
A. Interpretation	Redefinition/clarification	Redundancy
B. Retrieval	Stimulation	
	Focus/salience	Simple contrast?
C. Judgment	Constraint (normative & logical)	Simple contrast
D. Response selection		
1. Mapping		
2. Editing		
a. Consistency	Constraint (normative & logical)	Simple contrast?
	Focus/salience	
b. Self-presentation	Focus/salience	

First, sequence effects are hard to relate to the Tourangeau-Rasinski scheme. It is possible to associate them with the various steps, but because sequence effects are not related to substance, the carryover versus backfire distinction seems to apply well. For example, fatigue effects in general lead to less accurate and less thoughtful response. Retrieval would tend to be less thorough and less accurate, and judgments would tend to be less considered and more labile. These effects result from the number of prior questions but do not seem to represent either carryover or backfire effects as Tourangeau and Rasinski conceptualized them.¹⁴ The same would seem to apply to rapport, learning, and initial frame of reference effects.

Second, for substantive context effects, the matching of standard types to the Tourangeau-Rasinski scheme is more appropriate and useful. For the interpretation stage, there seem to be standard types that closely match both carryover and backfire effects. Redefinition/clarification effects are carryover interpretation effects whereby prior questions change or create meaning for a following question. Redundancy effects are backfire interpretation effects where a topic covered by a prior question is excluded from consideration in responding to a later question.

Carryover effects at the retrieval stage would seem to cover two types of traditional effects: stimulation effects and some, but not all, focus and general salience effects. Stimulation effects either result from more thorough memory searching or lead to overreporting through telescoping. Focus/salience effects at this stage come about from biased sampling of memory due to selective prim-

¹⁴Tourangeau and Rasinski (1988) explicitly admit that their framework is not comprehensive and separately discuss sequence effects.

ing. Backfire retrieval effects occur when respondents "discount or actively suppress information [created by prior questions] that they regard as suspect or irrelevant" (Tourangeau & Rasinski, 1988, p. 305). It is not clear if any of the standard categories are examples of this process, although some simple contrast effects may be of this type.

In the judgment stage, normative and logical constraint effects represent carryover effects, as do some of the more amorphous focus effects. For example, context may create or enhance in people's minds the norm of even-handedness as a standard for judging later questions. Simple contrast effects can represent backfire judgment effects where one's evaluation of a subsequent question is contrasted to an earlier standard of judgment.

The response selection stage covers some rather distinct processes that might well be thought of as involving different steps. The mapping process of figuring what response represents a respondent's attitude does not appear to be related to any of the standard types of context effects. (Learning effects would presumably lead to less error at this step.) At the editing stage, carryover effects are once again represented by both normative and logical constraint effects and more diffuse focus/salience effects. Whether a normative effect belongs in this category rather than in the judgment stage depends on whether the acknowledgment of the norm resulted from a perhaps unconscious and sincere application of a norm or the conscious and strategic decision to follow a norm in order to appear consistent. In the case of the even-handedness norm on the Communist/American reporters questions, a judgmental carryover effect exists if a norm is created in a respondent's mind that shapes attitudes toward a later question. A consistency response selection effect occurs if the norm does not really shape one's attitude toward allowing Communist reporters but one changes one's response in order to appear consistent with the norm. Although logically distinguishable, empirically separating these two types of normative effects would be difficult. Self-presentation/social desirability effects during editing involve related strategic responding and also cover some forms of focus/salience effects.

For backfire editing effects, there again do not appear to be any examples from the standard classification or survey literature. Tourangeau and Rasinski do describe a hypothetical example, which they call a "moderation effect." Consider the case of a person who has a self-perception as being a moderate on an issue (e.g., abortion). If that person has answered several questions in a pro-abortion direction, he or she might answer subsequent questions in an anti-abortion direction in order to maintain a moderate image, even though the respondent's true attitude on the individual subsequent questions might be pro-abortion.

What has the marrying of standard survey research classifications to the Tourangeau-Rasinski scheme suggested? First, the exclusion of sequence effects from their scheme indicates that their scheme is not comprehensive. Second, the lack of clear survey examples for several of their types suggest that these may be rare effects. Third, the splitting of several traditional effects across more than one of their categories shows that some existing types clearly involve different processes, and employing the Tourangeau-Rasinski framework clarifies the difference.

Fourth, the appearance within their categories of different standard effects (e.g., stimulation and focus among carryover retrieval effects) suggests that there are useful divisions within their stages (similar to those in the response selection stage). Fifth, the difficulty of empirically distinguishing between what stage or stages are involved indicates that great challenges await in identifying the processes that create context effects. Finally, combining elements of the standard classification scheme with the Tourangeau-Rasinski scheme will probably create a better framework than either alone.

Order effects can be induced by a variety of cognitive and social-psychological processes. These processes can intercede at various steps in the question-answering process and can impinge on later questions in at least two ways (carryover and backfire). Sometimes position alone is sufficient to create an effect, whereas other effects are stimulated by the substance of prior questions and often by a respondent's implicit or explicit attitude toward the prior substance.

In addition, many types of order effects can interact and commingle. For example, fatigue effects can be reduced or increased by question form and the topics covered. Although there may be a general fatigue curve associated with time or number of responses, the slope of this curve may be increased or decreased by such factors as the format of the questions and the interest and difficulty of the questions involved. In fact, two or more different (and even conflicting) effects may be relevant in the same instance. For example, extended discussions of a topic usually result in more interest in that issue being subsequently reported. Bishop (1987), however, found that when the discussion included several difficult knowledge questions about which most people lacked information, interest decreased. In this instance, it appears that the stimulation effect was overcome by a logical connection effect that linked low knowledge with low interest. It is probably such interactions that explain various failures to replicate order effects or to generalize to apparently similar circumstances (Schuman & Presser, 1981; Tourangeau & Rasinski, 1988; Turner & Martin, 1984).¹⁵ Order effects, alas, are not of Horatian simplicity.

Conclusion

Refining our understanding of order effects will not be an easy task given (a) the large number of different processes involved; (b) the difficulty of distinguishing between competing explanations; (c) the interaction of order effects with such other factors as question type (e.g., behavioral, affective), question specificity (T. W. Smith, 1988a), question vagueness (Turner & Martin, 1984; Zaller, 1988), question centrality (Turner & Martin, 1984; Krosnick & Schuman,

¹⁵As Tourangeau and Rasinski (1988) note, "Context effects are often unstable; this instability may reflect the number and complexity of the processes that are responsible for the effects, as well as the large number of variables that can influence the size and direction of the context effects" (p. 311).

1988), response type (substantive response vs. nonresponse), history (e.g., the Communist/American reporters and parental/student party identification—Willick & Ashley, 1971; Schuman & Presser, 1981; Hyman & Sheatsley, 1950), mode and pace of administration (Tourangeau & Rasinski, 1988), ambivalent or conflicted attitudes (Tourangeau & Rasinski, 1988), and other factors; and (d) interactions between different types of effects.

One key to further progress is simply to apply theoretical models, setting up experiments to test specific hypotheses about the causality of order effects and clearly choose between competing explanations. In particular, experiments need to be devised that can determine which stages in the question-answering process are involved. This will necessitate moving beyond simple split-ballot experiments. Useful as split ballots are with their experimental controls, we shall have to apply even more elaborate designs to gain a better understanding of the mental processes that cause order effects.

One useful approach is the think-aloud procedure, by which respondents are asked to relate their cognitive processes orally while these are occurring. Limitations are that it probably works best for conscious mental processes and that verbalization may significantly alter the mental process being employed.

A second promising approach would be the addition of a follow-up question after the antecedent and subsequent questions that would inquire about what the respondent was thinking about (Bishop, 1985, and chap. 11 of this volume; Tourangeau & Rasinski, 1986). Take Kalton's example (Kalton, Collins, & Brook, 1978), in which evaluations of driving standards were rated more positively immediately after a similar question about the driving standards of young drivers. Kalton et al. hypothesize that the more positive evaluation of drivers in general resulted from a subtraction effect that excluded young drivers from consideration in the second question. We should be able to test for this effect by asking after the general driving condition either an open-ended question about what type of driver one had in mind or a more focused closed question such as, "When you answered the question about general driving standards, were you thinking mostly about young drivers, middle-aged drivers, or older drivers?" If a subtraction effect were operating, there should be a reduction in references to young drivers when the general question was preceded by the question about young drivers. Other follow-up questions could be used to test the operation of other effects.

Another possibility is the use of questions probing other dimensions besides affect: importance, salience, information, knowledge, and commitment (Gallup, 1947; Schuman & Presser, 1981; T. W. Smith, 1981c). By learning with what dimensions and conditions order effects interact, we should better understand their causes (Tourangeau & Rasinski, 1986). Similarly, attributes of the questions, such as vagueness and response categories, could be explored.

Another useful approach would be a test/retest design in which four orders could be used (A1B1A2B2, A1B1B2A2, B1A1B2A2, and B1A1A2B2). This would allow a comparison of the consistency of each item in each order (Hayes, 1964; T. W. Smith & Stephenson, 1979). Given certain assumptions, it would

also permit an intrarespondent analysis of order effects. Alternatively, one might ask respondents the subsequent question later in the same interview in a different context. Interviewers could then reconcile discrepancies in responses. Through these and other elaborations of the basic split-ballot technique, it should be possible to examine directly the causes of context effects and gain a deeper understanding of the mental processes involved.

By more fully analyzing split-ballot order experiments, by elaborating these experiments with specific inquiries about mental processes and other auxiliary items, and by greater grounding in appropriate cognitive and social-psychological theories, we should be able to advance our understanding of order effects. Although the natural complexity of language and human cognition will undoubtedly hinder precise and comprehensive generalizations about order effects, thorough and cumulative analysis of sophisticated theory-driven experiments should greatly advance the art of ordering questions.

Acknowledgments. Parts of this chapter are adapted from T. W. Smith (1986a). This research was done for the General Social Survey project directed by James A. Davis. This project is supported by the National Science Foundation, Grant No. SES-87-18467.