COGNITIVE PSYCHOLOGY 24, 1–55 (1992)

Order Effects in Belief Updating: The Belief-Adjustment Model

ROBIN M. HOGARTH AND HILLEL J. EINHORN

University of Chicago

Much literature attests to the existence of order effects in the updating of beliefs. However, under what conditions do primacy, recency, or no order effects occur? This paper presents a theory of belief updating that explicitly accounts for order-effect phenomena as arising from the interaction of information-processing strategies and task characteristics. Key task variables identified are complexity of the stimuli, length of the series of evidence items, and response mode (Step-by-Step or End-of-Sequence). A general anchoring-and-adjustment model of belief updating is proposed. This has two forms depending on whether information is *processed* in a Step-by-Step or End-of-Sequence manner. In addition, the model specifies that evidence can be encoded in two ways, either as a deviation relative to the size of the preceding anchor or as positive or negative vis-à-vis the hypothesis under consideration. Whereas the former (labeled estimation mode) results in data consistent with averaging models of judgment, the latter (labeled evaluation mode) implies adding models. Conditions are specified under which (a) evidence is encoded in estimation or evaluation modes and (b) use is made of the Step-by-Step or End-of-Sequence processing strategies. The theory is shown both to account for much existing data and to make novel predictions for combinations of task characteristics where current data are sparse. Some of these predictions are examined and validated in a series of five experiments. Finally, both the theory and the experimental results are discussed with respect to the structure of models of updating processes, limitations and extensions of the present work, and the importance of developing a procedural theory of judgment.

How do people update beliefs across time? Imagine, for example, a physician who is uncertain whether a patient is suffering from a particular disease and orders several tests. Knowledge of the test results can change the physician's level of uncertainty as well as the therapeutic alternatives

This work was supported by a contract with the Office of Naval Research as well as research funds from The University of Chicago, Graduate School of Business. The authors gratefully acknowledge the able research assistance of Haim Mano, Ann McGill, John Hand, Jay Koehler, Brian Gibbs, Craig McKenzie, and Alison Marquis as well as useful comments by Paul Andreassen, Alison Ashton, Robert Ashton, William Goldstein, Stephen Hoch, Irwin Levin, Charles Judd, George Loewenstein, William Messier, Richard Tubbs, Elke Weber, and several anonymous reviewers. The present paper also owes much to penetrating critiques of earlier versions by Norman H. Anderson and Joshua Klayman.
Address correspondence and reprint requests to R. M. Hogarth, University of Chicago, Graduate School of Business, 1101 East 58th Street, Chicago, IL 60637.

0010-0285/92 $7.50

considered. Alternatively, imagine that a new acquaintance has made a favorable impression on you. Subsequent interactions, however, may lead you to revise this opinion.

A critical feature of belief updating is its sequential nature. As stated by Anderson (1981),

> In everyday life, information integration is a sequential process. Information is received a piece at a time and integrated into a continuously evolving impression. Each such impression, be it of a theoretical issue, another person, or a social organization, grows and changes over the course of time. At any point in time, therefore, the current impression looks both forward and back. (p. 144)

This paper presents a descriptive theory of belief updating that can be applied to many substantive domains. The focus is on order effects. In the paradigmatic situation considered here, a person responds to the presentation of several pieces of evidence by expressing an opinion about a specific proposition or hypothesis. However, were the person to process the same information in a different order, would this affect the final judgment? More specifically, under what conditions does information processed early in the sequence have greater influence, i.e., produce a primacy effect? Under what conditions is later information more important, i.e., a recency effect? Under what conditions is order irrelevant?

The paper is organized as follows. We first briefly summarize experimental findings on order effects and discuss different hypotheses that have been advanced to explain their existence. In particular, we show how order effects are sensitive to both procedural and contextual variables. Second, we present our theory of updating which is based on a general anchoring-and-adjustment model. This model, however, can take several forms depending on subprocesses that govern (a) whether evidence is encoded as a deviation relative to the preceding anchor or as positive or negative vis-à-vis the hypothesis under consideration, (b) whether beliefs are revised in response to each piece of evidence or only after all the evidence has been processed, and (c) the adjustment process. We further specify the conditions under which these subprocesses are used. Although our model presumes a simple information-processing strategy, its interaction with different task variables yields a complex pattern of predictions concerning order effects. These predictions are validated in two ways: first, by reference to the existing experimental literature and, second, by testing the model's predictions for cases where experimental evidence is sparse. We conclude by discussing the implications of our theory and results for understanding the nature of belief updating and order effects. At a more general level, our analysis of order-effect phenomena can be seen as an illustration of how the observed complexity of behavior can arise through the interaction of simple psychological processes with the demands of an infinitely varied environment (Brunswik, 1952; Simon & Newell, 1971; Hogarth, 1986).

A BRIEF REVIEW OF ORDER EFFECTS IN BELIEF UPDATING

Belief updating is a ubiquitous human activity. It is an essential component in such diverse areas as probabilistic inference (Peterson & Beach, 1967; Edwards, 1968; Slovic & Lichtenstein, 1971; Hogarth, 1975; Schum, 1980; Fischhoff & Beyth-Marom, 1983), decision theory (Raiffa & Schlaifer, 1961; Winkler, 1972), economics (Camerer, 1987), social cognition (Fishbein & Ajzen, 1975; Nisbett & Ross, 1980; Anderson, 1981; Hastie, 1983), jury decision making (Schum & Martin, 1982; Davis, 1984; Pennington & Hastie, 1986), communication and persuasion (Hovland, Janis, & Kelley, 1953), attitude change (Triandis, 1971; Cooper & Croyle, 1984; Petty & Cacioppo, 1986), causal inference (Jones, 1979; Einhorn & Hogarth, 1986; Carlson & Dulany, 1988), and psychophysics (Green & Swets, 1966).

Although the study of order effects has played an important role in this literature, there is disagreement concerning what kinds of order effects are most prevalent. For example, in an influential book on social judgment, Nisbett and Ross (1980) stated, "Although order of presentation of information sometimes has no effect on final judgment, and recency effects sometimes are found, these are the exception; several decades of psychological research have shown that primacy effects are overwhelmingly more probable" (p. 172). This conclusion, however, is contradicted by Davis (1984) in a review of studies of jury decision making that indicates greater prevalence of recency effects as well as work by Anderson and his colleagues whose studies have demonstrated both primacy and recency (see Anderson, 1981).

In the present work, we consider order effects of the following type: There are two pieces of evidence, A and B. Some subjects express an opinion after seeing the information in the order A-B; others receive the information in the order B-A. An order effect occurs when opinions after A-B differ from those after B-A. (On occasion, within-subject analysis is also possible; see, e.g., Shanteau, 1970.)

This definition of order effects distinguishes our topic from others in which the terms primacy and recency have been used. For example, Pennington and Hastie (1986) have demonstrated the importance of the first information jurors receive when constructing a mental representation of a purported crime—a "primacy" effect. In other situations, belief revision is compared with a normative standard such as Bayes' theorem (see, e.g., Peterson & DuCharme, 1967). Third, results obtained in the belief perseverance paradigm (Ross & Lepper, 1980) have been cited as favoring primacy (Nisbett & Ross, 1980, Chap. 8). However, the effects reported in these studies do not qualify as order effects by our definition because the same information is not presented to different groups of subjects in different orders.

The order-effect studies we consider within the paradigm defined above involve a wide range of tasks including impression formation, probability estimation, assessment of guilt or innocence in mock trials, attributions of performance, estimates of contingencies, and judgments of weights.

Task Differences

Studies of order-effect phenomena can be classified on many dimensions, varying, for instance, from the semantic content of stimuli (e.g., affective vs evaluative judgments), to specific differences in procedures. The task variables we consider are the complexity of the individual items of evidence to be processed, the length of the series of items, and response mode. We first elaborate on each of these variables.

Complexity. It is well known that information-processing strategies reflect the complexity of task requirements. In studies of choice, for example, as complexity increases people cannot use the more comprehensive strategies that they may apply to simpler choice problems and, instead, resort to strategies that ease cognitive strain (see, e.g., Payne, Bettman, & Johnson, 1990). It is reasonable to postulate that differential task complexity will affect strategies people use for processing information in belief-updating tasks. In this paper we define complexity operationally with respect to characteristics of the individual pieces of evidence that are to be integrated. Specifically, complexity is assumed to be an increasing function of both the amount of information that needs to be processed for each piece of evidence and lack of familiarity with the task. We justify the former by appealing to the well-known limitations on human information-processing abilities and the latter by the fact that an important component of expertise in a specific domain involves strategies for coping more effectively with greater amounts of information (Newell & Simon, 1972).

Length of series. Apart from the amount of information to be processed for each piece of evidence, the number of pieces to be evaluated can vary. As the number of pieces increases, we expect two kinds of effect. First, subjects can tire if asked to process many pieces of information. Second, as information accumulates, beliefs are expected to become less sensitive to the impact of new information because this represents an increasingly small proportion of the evidence already processed.

Response mode. Many studies have shown that judgments are sensitive to the manner in which they are elicited (Einhorn & Hogarth, 1981; Hogarth, 1982; Goldstein & Einhorn, 1987). Our third task variable concerns two response modes that have been commonly employed in the literature.

TABLE 1
Classification of Results of Order Effect Studies According to Task Characteristics[a]

Evidence items:	Simple		Complex		
Response mode:	EoS	SbS	EoS	SbS	Total
		Short series			
Primacy	19	—	1	—	20
Recency	5	16	7	2	30
No effect	3	—	1	—	4
		Long series			
Primacy	12	2	2	—	16
Recency	2	—	1	2	5
No effect	—	—	1	—	1

[a] Table entries are studies detailed in Appendix A.

These are the Step-by-Step procedure in which subjects are asked to express their beliefs after integrating each piece of evidence in a given sequence, and the End-of-Sequence procedure where subjects only report their opinions once all the information has been presented. These procedures will be henceforth abbreviated as SbS and EoS, respectively.

Table 1 provides a summary of our classification of the outcomes of the studies we examined using the three task variables discussed above and indicates the frequency with which different types of order effects have been observed for different combinations of the task variables.[1] The details underlying Table 1 are to be found in Appendix A. However, before commenting on Table 1, we briefly discuss the operational criteria used for classifying the studies.

For complexity, we adopted two classes: "Simple" and "Complex." Many studies could be unequivocally allocated to the Simple category because the information subjects were required to process involved only a single item for each piece of evidence (e.g., trait adjective or number) in tasks with which one could reasonably assume familiarity, e.g., forming impressions of people or estimating the average of several numbers. Evidence items were judged to be Complex if they involved a large amount of information (e.g., 600-word messages, Crano, 1977) or unfamiliar stim-

[1] In searching the literature, we followed up on references contained in the reviews by Anderson (1981), Nisbett & Ross (1980), and Slovic & Lichtenstein (1971), as well as consulting *Psychological Abstracts* for studies of order effects in experiments on judgment. In addition, several colleagues brought particular studies to our attention. The fact that order effects are not the primary focus of investigation in many experiments means that it is difficult to identify all studies. We therefore do not claim that the basis for our review is comprehensive. On the other hand, we do believe it is representative of the universe of relevant studies.

uli (e.g., estimating averages of sets of noises, Parducci, Thaler, & Anderson, 1968, or weights on a pulley, Anderson & Jacobson, 1968).[2]

Categorizing studies by whether the series of items of evidence to be integrated was "Short" or "Long" was made easy by the fact that studies fell conveniently into two groups, the Short with between 2 and 12 items, and the Long with 17 or more. Finally, classification by response mode, i.e., SbS or EoS, was straightforward.

Table 1 is organized in two sections; one dealing with tasks involving Short series of evidence items, the other with Long series. Each entry in the table refers to one published study even though these studies might contain more than one experiment. The exception to this rule is provided by cases where studies examined the effects of different combinations of the three task characteristics we had identified as relevant. The outcomes of the different task combinations within such studies are each represented as data points in Table 1. There are thus 76 data points for the 60 studies detailed in Appendix A.

A simple examination or "head count" of the outcomes of these studies reveals sharp differences that depend on the classifications adopted. Note, first, that the total column on the right-hand side of Table 1 documents the presence of primacy and recency effects for both Short and Long series. The majority of studies (43 of the 76 data points) have examined Simple evidence items for Short series. These have typically involved between, say, three and six evidence items such as trait adjectives to be used for making social judgments of "likableness" (e.g., Anderson & Barrios, 1961) or simple (e.g., two-digit) numbers to estimate an average (e.g., Hendrick & Constantini, 1970b). These tasks reveal a clear pattern: primacy in 19 of 27 EoS studies and recency in 16 of 16 SbS studies. Moreover, it is significant that the EoS studies resulting in recency involved experimental manipulations that probably affected the judgment process. In Anderson (1968) and Hendrick and Constantini (1970a), for example, subjects were required to pronounce each piece of information (adjectives in impression formation tasks); in Luchins (1957b) recency occurred when another task was interpolated between two blocks of pieces of evidence; and whereas 12 independent trait adjectives re-

[2] Difficulties were sometimes experienced in judging whether evidence items were simple or complex. For example, evidence items in Dreben, Fiske, & Hastie (1979) involved two sentences totaling about 23 words. We chose to classify this as simple because subjects were only given 10 s to read the words and provide ratings (and presumably were able to do so). In Luchins (1957a, 1957b), subjects were presented with short paragraphs describing individuals. We chose to classify these studies as simple because the paragraphs contained a number of different traits (or evidence items) that were embedded in the text. As can be seen from Table 1, the main conclusions reached by our analysis would not be substantially affected if a few such studies were reclassified.

sulted in the modal finding of primacy in Strange, Schwei, and Geiselman (1978), recency occurred when the same 12 adjectives were characterized as 4 pieces of evidence each consisting of 3 trait adjectives. There are fewer data concerning Short, Complex tasks. However, the effect of response mode does not seem to be the same as for Short, Simple tasks. Almost all (9 of 11) have demonstrated recency independent of response mode. Among tasks involving Long series, 14 of 16 studies in the Simple category resulted in primacy, a finding that appears to be independent of response mode. On the other hand, data for studies classified as Complex are inconclusive.

To summarize, our task analysis of order effects leads to the following conclusions: (1) response mode makes a difference in the case of Short, Simple tasks. EoS induces primacy, SbS induces recency; (2) primacy seems to obtain when tasks are Simple but Long (this is also independent of response mode); and (3) recency is associated with more complex tasks (independent of response mode).

It is important to note that this analysis of the literature does not reflect two task variables that are important to the predictions of our model presented below. These are, first, how task demands affect whether evidence is encoded relative to constant or variable reference points (the preceding anchor) and, second, whether the pieces of evidence to be aggregated are consistent or inconsistent. Few, if any, studies have explicitly considered how evidence is encoded (but see Lopes, 1985, 1987a), and almost all studies reviewed involved inconsistent information (however, see Anderson, 1968; Chalmers, 1971).

Theories. In accounting for the results discussed above, most theoretical attention has been given to tasks that we have described as Short and Simple. In his 1946 study of impression formation, Asch suggested that the first piece of information encountered when forming an impression of another person set up a direction that, in some cases, might change the interpretation of subsequent information. Thus, someone described as "intelligent—tall—mean" would be judged more favorably than someone described as "mean—tall—intelligent." That is, "intelligent" is taken as a positive attribute in the first instance but less so when following "mean." An alternative explanation of the same phenomenon follows from the notion that when someone describes another person, there is a natural presumption that order reflects the importance of the information provided.

Working within the framework of information integration theory, these views were challenged by Anderson (1981) who showed that neither "change of meaning" (as proposed by Asch), nor the "natural presumption hypothesis," nor "inconsistency discounting" (whereby evidence that is inconsistent with what preceded it is deemphasized) is necessary to induce primacy effects when subjects are required to respond to an EoS response mode. Instead, Anderson claimed that primacy can be explained by a process of "attention decrement" whereby people pay less attention to successive items of evidence. This hypothesis is supported by the finding that primacy is replaced by recency in the presence of an SbS response mode (or, as noted above, when the judgment process is affected by demands such as requiring subjects to pronounce each trait aloud). To describe the belief-updating process, Anderson used a serial integration model which represents updating judgments in the form of weighted averages of the scale values of the evidence items where the inferred weights reflect the effects of serial position (see Anderson, 1981).

The attention-decrement explanation, however, raises two important issues. First, what is the domain to which the hypothesis applies? (See also Jones & Goethals, 1972.) For example, in a task involving the estimation of the average lengths of lines, Weiss and Anderson (1969) reported recency effects with both EoS and SbS response modes. In addition, recency was found in two separate experiments using the same complex evidence items (paragraphs describing U.S. presidents) but different response modes, SbS in Anderson and Farkas (1973) and EoS in Anderson (1973a). And second, whereas attention decrement provides an explanation for primacy, what causes recency?

THE BELIEF-ADJUSTMENT MODEL

Our theory assumes that people handle belief-updating tasks by a general, sequential anchoring-and-adjustment process in which current opinion, or the anchor, is adjusted by the impact of succeeding pieces of evidence.

In algebraic terms, the model can be written

$$S_k = S_{k-1} + w_k[s(x_k) - R], \tag{1}$$

where

S_k = degree of belief in some hypothesis, impression or attitude after evaluating k pieces of evidence ($0 \leq S_k \leq 1$).

S_{k-1} = anchor or prior opinion. The initial strength of belief is denoted S_0.

$s(x_k)$ = subjective evaluation of the kth piece of evidence. (Different people may accord the same evidence, x_k, different evaluations. Bounds on $s(x_k)$ are specified below.)

R = the reference point or background against which the impact of the kth piece of evidence is evaluated.

w_k = the adjustment weight for the kth piece of evidence ($0 \leq w_k \leq 1$).

Within this general model there are three important subprocesses which concern (a) *how evidence is encoded*—relative to constant or variable reference points, (b) *how evidence is processed*—whether beliefs are revised in response to each piece of evidence or only after all the evidence has been processed, and (c) *how the adjustment is accomplished.* In this section, we specify both these subprocesses and our hypotheses concerning the conditions under which they are used. In the following section, we present the implications of the model for order-effect phenomena.

Encoding

In recent years there has been growing awareness of the effects of encoding on outcomes of judgment and choice (Kahneman & Tversky, 1979; Lopes, 1987b). This is also important in belief updating where we hypothesize that, prior to actual updating, evidence is evaluated relative to a reference point, R. We further argue that R will be equal either to the prior anchor, S_{k-1}, or a constant. In other words, evidence is encoded either relative to the level of current belief or in an absolute manner. This therefore raises the need to specify both the implications of these alternative reference points or backgrounds and when they are adopted.

We distinguish between *evaluation* and *estimation* tasks. In evaluation tasks, people encode evidence as positive or negative relative to the hypothesis under consideration. Imagine, for example, assessing the validity of a theory or causal hypothesis in terms of whether it is true where, conceptually, your current belief can be expressed by some value on the continuum between "false" (=0) and "true" (=1). Irrespective of the level of current belief, supporting evidence increases belief in the theory whereas disconfirming evidence decreases it. In evaluation tasks, evidence is seen as bipolar (cf. Lopes, 1982) relative to the hypothesis (confirming versus disconfirming) such that $-1 \leq s(x_k) \leq +1$ and $R = 0$ (cf. Einhorn & Hogarth, 1986; Carlson & Dulany, 1988). In this case Eq. (1) takes the form

$$S_k = S_{k-1} + w_k s(x_k). \tag{2}$$

In contrast, estimation tasks involve assessing some kind of "moving average" (e.g., impression of "likableness") that reflects the position of each new piece of evidence relative to current opinion. Imagine, for example, assessing how much you like another person. Conceptually, current opinion can be thought as being expressed on a scale from "dislike" (=0) to "like a lot" (=1). New evidence is seen as another measure of the likableness of the person. It will therefore increase current opinion if greater than the latter, but decrease it if smaller. In other words, in estimation tasks people are sensitive to the difference between the location of the current anchor (i.e., S_{k-1}) and the level of opinion suggested by the evidence to be integrated. As an estimate or measure of opinion, evidence is seen as unipolar (cf. Lopes, 1982) such that $0 \leq s(x_k) \leq 1$ and $R = S_{k-1}$. In this case Eq. (1) can be written

$$S_k = S_{k-1} + w_k[s(x_k) - S_{k-1}], \tag{3}$$

which, after rearranging terms, leads to the averaging form

$$S_k = (1 - w_k)S_{k-1} + w_k s(x_k). \tag{4}$$

It is important to note that whereas both Eqs. 2 and 3 represent anchoring-and-adjustment strategies, the data produced by the two subprocesses will have different patterns. As indicated by Eq. (4), estimation tasks will result in data that fit averaging models; data from evaluation tasks, however, will fit adding models.

To illustrate further qualitative differences between the evaluation and the estimation modes, consider a task in which subjects are required to update opinions on the basis of two pieces of evidence in a step-by-step manner. Assume further that both pieces of evidence are "positive" but that one is strong and the other weak. To be specific, let the evidence items have values of +.7 and +.3, respectively, if assessed on the bipolar evaluation scale, and values, say, of .9 and .6 if assessed on the unipolar estimation scale; in addition, assume that current opinion is .5. Now consider the implications of processing the information in both the weak–strong and the strong–weak orders.

In evaluation mode, the responses to both pieces of evidence will involve upward revisions of belief whether the information is considered in the weak–strong or strong–weak orders. This is because the evidence has been encoded as positive. In the estimation mode, responses to both pieces of evidence will also involve upward revisions in the weak–strong order, i.e., because .6 > .5, and .9 > "average" of .6 and .5. However, this will not be the case in the strong–weak order. Here, the response to the strong piece of evidence will be an upward adjustment (.9 > .5), but the effect of the weak evidence will be a downward adjustment, i.e., because .6 < "average" of .9 and .5. In other words, whereas the qualitative patterns of responses to data in the weak–strong order are similar for the evaluation and estimation modes, they differ in the strong–weak order.

This distinction between estimation and evaluation modes was nicely illustrated by Lopes (1985, 1987a) in studies investigating why people sometimes revise probabilistic beliefs in a direction opposite to that prescribed by Bayes' theorem. In an order-effect paradigm similar to the hypothetical situation described above, Lopes' subjects made downward revisions of belief in response to weak evidence received in the strong–weak order as opposed to the upward revisions prescribed by Bayes' theorem, which, it should be recalled, has an adding formulation. (See also Shanteau, 1970, 1972.) Lopes further showed that subjects could be instructed to adopt a judgmental strategy that mitigated against errors vis-à-vis the Bayesian model. From our viewpoint, it is significant that this involved teaching subjects to start by labeling or evaluating the evidence in terms of which of two hypotheses it favored prior to adjusting belief in the direction of the label. In other words, subjects were taught to use the evaluation as opposed to estimation mode.

Both Lopes' studies and the extensive evidence of averaging processes in judgment (see, e.g., Anderson, 1981) suggest that estimation may be more frequently used than evaluation. In addition, many task instructions in the literature have implicitly or explicitly required subjects to use estimation by requiring the assessment of "averages." When, however, do people use evaluation as opposed to estimation in the absence of explicit instructions? We believe two considerations are relevant.

First, evaluation and estimation differ in the kinds of implicit scales involved in the judgment process. The outcomes of evaluation tasks involve judgments on dichotomous scales of the "true–false" variety as opposed to assessments of "how much" that are typical of estimation tasks. This distinction is further emphasized by the scales used for assessing evidence in the two kinds of tasks. In evaluation, evidence is conceptually measured on a bipolar scale ($-1 \leq s(x_k) \leq +1$) but then transformed onto a unipolar scale ($0 \leq S_k \leq 1$). Estimation, on the other hand, involves the transformation of evidence from one unipolar scale ($0 \leq s(x_k) \leq 1$) onto another ($0 \leq S_k \leq 1$).

Second are considerations of cognitive economy and the extent to which the form of the evidence is more compatible with estimation or evaluation (cf. Tversky, Sattath, & Slovic, 1988). Lopes, for example, has speculated that subjects tend to use evaluation-mode strategies when "stimuli are explicitly 'marked' for one hypothesis or another" (Lopes, 1985, p. 512) thereby facilitating a directional (positive or negative) interpretation of the evidence irrespective of current opinion. Similar effects can be expected when evidence is presented in a form that requires explicit interpretation as to whether it is for or against the hypothesis. For example, imagine having to evaluate testimony in a court case that requires a verdict of guilt or innocence. Here the hypothesis is bipolar in nature and evidence must be explicitly interpreted as to whether it incriminates or exonerates the accused. In contrast, consider a situation where, starting with a given level of belief in, say, guilt, one is told that the evidence of three of four witnesses also favors a guilty verdict. In this case, the form of the information requires little interpretation and is implicitly scaled in a unipolar manner compatible with estimation; i.e., how large is ¾ relative to one's current position?

In summary, we hypothesize that the use of evaluation as opposed to estimation will primarily depend on both the nature of the opinion to be expressed (dichotomous belief versus estimates on a continuous scale) and the extent to which the evidence is more compatible with estimation (on a unipolar scale) or evaluation (on a bipolar scale). In Experiment 5 below we test the implications of the hypothesized differences between evaluation and estimation.

Processing

So far, we have discussed the model as though $s(x_k)$ represented a single piece of evidence. However, by also allowing $[s(x_k) - R]$ to represent the net impact of several pieces of information, we can use the model to illuminate two important variants of Eq. (1). These are the Step-by-Step (SbS) and End-of-Sequence (EoS) *processes* where these terms have meanings similar to the two response mode conditions discussed above. Specifically, when using an SbS *process,* a person is assumed to adjust his or her opinion incrementally by each piece of evidence processed. With an EoS *process,* on the other hand, the initial anchor is adjusted by the aggregate impact of the succeeding set of evidence. To illustrate this distinction, imagine forming an impression of "likableness" based on a series of trait adjectives such as "intelligent—tall—mean." In using an SbS process, a person is assumed to anchor on "intelligent" and then to update this impression incrementally, first by "tall," then by "mean." Using an EoS process, only one adjustment is made. The impression is anchored on the first or first few pieces of evidence and then adjusted by the net effect of the others. If, for example, "intelligent" determines the anchor, the adjustment results from the impact of the net aggregate impression of the words "tall" and "mean."

Mathematically, the EoS anchoring-and-adjustment strategy can be represented by

$$S_k = S_0 + w_k[s(x_1, \ldots, x_k) - R], \tag{5}$$

where $s(x_1, \ldots, x_k)$ is some function, possibly weighted average, of the individual subjective evaluations (or scale values) of the items of evidence that follow the anchor.

It would be convenient to assume that people always use an SbS process when faced with an SbS response mode and an EoS process when faced with an EoS response mode. However, whereas we are confident in assuming the former, the latter is more complicated. Consider Fig. 1 which shows the four cells that arise from crossing the two forms of process with the two forms of response mode. Note, first, that when a task demands an SbS response mode, one cannot use an EoS process. In other words, an SbS response mode necessarily invokes an SbS process.

		Response mode	
		SbS	EoS
Process	SbS	All tasks	Complex evidence items and/or long series
	EoS	Impossible	Simple evidence items and short series

FIG. 1. Compatibility between SbS and EoS *processes* and *response modes*.

Second, an EoS response mode can be met by using either an EoS or SbS process where the latter involves making step-by-step revisions in beliefs but overtly verbalizing only the final opinion. Faced with an EoS response mode, when do people use an EoS or SbS process?

An important difference between the EoS and the SbS processes is the nature of the demands they make on memory and information-processing load. When faced, for example, with a sequence of four items of evidence, the EoS strategy requires aggregating the latter items prior to integrating them with the anchor. Aggregation, however, can be costly in terms of mental resources. In contrast, the step-by-step integration of each item of evidence in the SbS process makes minimal demands on memory and information-processing load. This difference suggests that the choice between EoS and SbS strategies is determined by the effects of task characteristics on cognitive capacity. We therefore assume that, when required to provide EoS responses, people are more likely to use an SbS process as the relative complexity and/or length of the series of evidence items increases. As information-processing demands increase, people are forced into using the SbS strategy in order to cope with the cognitive demands of the task (i.e., the upper right cell of Fig. 1).

In general, we assume that people try to match cognitive strategy with response mode but shift strategies if this proves too demanding (cf. Payne et al., 1990). The choice between EoS and SbS processes can therefore be summarized as follows (see also Fig. 1):

1. When the response mode is SbS, the SbS process model is always used.

2. When the response mode is EoS:

(a) the EoS process model is used for short series of cognitively simple evidence items; and

(b) the SbS process is used for cognitively complex evidence items and/or longer series in order to handle the information-processing demands of the task.

Adjustment

The underlying purpose of adjusting beliefs to the impact of new evidence is adaptation. Thus the adjustment weight, w_k, should depend on both the sign of the impact of the evidence, i.e., of $[s(x_k) - R]$, and the level of the anchor, S_{k-1}. For $s(x_k) \leq R$, we argue that w_k is proportional to S_{k-1}. The rationale is the following: Imagine that your current position (i.e., anchor) is weak and a strong piece of negative evidence is received. Since your current position is already low, the new information cannot reduce the anchor a great deal (in absolute terms). Now consider the effect of the same negative evidence if your position were strongly held. We argue that the reduction of strength will be larger in the latter case. Note that this assumption implies a "contrast" effect since it says that large anchors are "hurt" more than smaller ones (given the same negative evidence). To borrow a boxing metaphor, the contrast effect implies that the bigger the anchor, the harder it will fall. Thus

$$w_k = \alpha S_{k-1}, \text{ when } s(x_k) \leq R. \tag{6a}$$

By an analogous argument, we assume that when $s(x_k) > R$, w_k is inversely proportional to the strength of the anchor, i.e.,

$$w_k = \beta\,(1 - S_{k-1}) \text{ when } s(x_k) > R, \tag{6b}$$

where α and β are constants.[3] To illustrate the effect of these assumptions, substitute Eqs. (6a) and (6b) into Eq. (1) so that the SbS process model can be written as

$$S_k = S_{k-1} + \alpha S_{k-1}[s(x_k) - R] \qquad \text{for } s(x_k) \leq R \tag{7a}$$

and

$$S_k = S_{k-1} + \beta(1 - S_{k-1})[s(x_k) - R] \qquad \text{for } s(x_k) > R. \tag{7b}$$

It is of interest to note that Eq. (7b) leads to a formulation identical to one proposed by James Bernoulli for combining the probabilistic impact of two pieces of positive evidence (see Shafer, 1976, pp. 75–77).

Whereas adaptation implies adjusting opinion for the impact of new

[3] For ease of exposition, we have chosen to work with these simple, tractable, functional forms. However, several other functional forms could also capture the qualitative predictions implied by our model.

evidence, one would not necessarily expect people to adapt to evidence at the same rate nor that the rate of adaptation be independent of the sign of the adjustment. We therefore define α and β ($0 \leq \alpha, \beta \leq 1$) to represent sensitivity toward "negative" and "positive" evidence, respectively (where the sign of the evidence is evaluated relative to R). Small values of α and β imply low sensitivity to new information; large values of α and β imply high sensitivity. We also posit that α and β are functions of both individual and situational variables. Some people, for example, may have a general tendency to weight negative information more heavily than positive. In other cases, a person may have such an investment in a particular belief that α is much smaller than β. (Consider a scientist who does not want to hear negative evidence about a favorite theory.) In general, one would expect that, in the absence of prior information or biases, a person would have large values of both α and β (i.e., be highly sensitive to evidence). However, as information accumulates and as people become more firmly committed to their beliefs, values of α and β would decline in a long series of evidence items.

Summary

Our general model is an anchoring-and-adjustment process. This involves three subprocesses that specify: (a) how evidence is *encoded*—relative to a variable or constant background, i.e., $R = S_{k-1}$ or $R = 0$; (b) how evidence is *processed*—Step-by-Step or End-of-Sequence, see Eqs. (1) and (5), respectively; and (c) the nature of the *adjustment* process. This embodies the contrast assumption and different forms depend on the sign of $[s(x_k) - R]$; see Eqs. (6a) and (6b). We further specified the conditions under which alternative forms of the subprocesses are used. We now examine the implications of our model for order effects.

ORDER-EFFECT PREDICTIONS

We first consider the order-effect predictions for the SbS and EoS subprocesses prior to showing how these interact with different task characteristics.

The SbS Process

Predictions for this subprocess depend on whether encoding involves the estimation or evaluation modes, i.e., $R = S_{k-1}$ or $R = 0$, respectively.

$R = S_{k-1}$

This leads to the process modeled by Eq. (3) which has exactly the same form as a model originally proposed by Anderson and Hovland (1957)—see also Anderson (1959). It has two important properties. First, as shown in Eq. (4) it implies an averaging process where S_k is a weighted average of the anchor, S_{k-1}, and the evidence, $s(x_k)$. Second, when $R = S_{k-1}$, the SbS process always predicts recency (for $\alpha,\beta \neq 0$)—see Appendix B.

R = 0

In this case, the SbS process no longer has a weighted average formulation and is similar in form to a model proposed by Carlson and Dulany (1988). Moreover, order-effect predictions depend on whether the sequence of evidence processed involves consistent or mixed evidence. By the former we mean sequences where the evidence encountered is either all negative or all positive; by the latter we mean sequences involving mixtures of positive and negative evidence.

When $R = 0$, the SbS process predicts no order effects for consistent evidence but always predicts recency in the sequential evaluation of mixed evidence except if α or β is zero (i.e., either negative or positive evidence is completely ignored). Furthermore, recency effects will be largest when both positive and negative evidence are strong and there is high sensitivity to evidence. These conditions are derived in Appendix C.

The EoS Process

As shown by Eq. (5), the EoS process is characterized by a single adjustment that reflects the *net aggregate* impact of the remaining pieces of evidence on the initial anchor. In fact, in many EoS experiments reported in the literature (see, e.g., Anderson, 1981), subjects have no explicit initial opinion such that the anchor is derived from the first piece(s) of evidence. In these circumstances, the structure of the EoS strategy contains within it a force toward primacy. Moreover, this result holds whether the evidence is all positive, all negative, or mixed. This occurs because, in the absence of an explicit starting point, we assume that the first piece of evidence, or an amalgamation of the first few pieces, serves as the anchor. Thus, to illustrate by denoting the first piece of evidence as the anchor, Eq. (5) can be rewritten as

$$S_k = s(x_1) + w_k[s(x_2, \ldots, x_k) - R], \tag{8}$$

where for a wide range of functions $s(x_2, \ldots, x_k)$, it follows that the effective weight accorded to $s(x_1)$ must be greater than the weight attached to any of the other pieces of evidence.

Subprocesses and Task Characteristics

The order-effect predictions that result from the intersection of our assumptions concerning subprocesses *and* task variables are summarized in Table 2. Note that in accordance with the preceding discussion, the

TABLE 2
Summary of Order-Effect Predictions

Encoding:	$R = S_{k-1}$		$R = 0$			
Type of evidence:	All		Mixed		Consistent	
Response mode:	EoS	SbS	EoS	SbS	EoS	SbS
Short series						
Simple	Primacy	Recency	Primacy	Recency	Primacy	No effect
Complex	Recency	Recency	Recency	Recency	No effect	No effect
Long series	Force toward primacy	Force toward primacy	Force toward primacy	Force toward primacy	Primacy	Primacy

predictions assume that the SbS response mode always invokes the SbS process, but that the EoS process will only be used when the EoS response mode is required in tasks involving short series of cognitively simple evidence items (i.e., the first row of Table 2). Second, the predictions reflect the same task variables used in the analysis summarized in Table 1 with the exceptions that we have now also distinguished between, first, whether evidence is encoded relative to S_{k-1} or 0 and, second, the distinction between series consisting of mixed and consistent evidence.

It is instructive to compare the predictions in Table 2 with the analysis of order effects summarized in Table 1.

Consider the first row of Table 2 which deals with tasks involving short series of simple evidence items. For EoS processing, the model always predicts primacy and this is the predominant outcome in the experimental literature when EoS response mode is required (19 of 27 studies, Table 1). For the SbS response mode, the evidence favors recency exclusively (all 16 studies, Table 1) and this is predicted by the model for all but the case of evaluation ($R = 0$) with consistent evidence. Here no effects are predicted. However, it is unclear whether any previous studies belong in this cell of the table.

The second row of Table 2 deals with short series of complex evidence items. Here the model predicts recency irrespective of response mode (i.e., SbS or EoS) except for the no-effects predictions for evaluation ($R = 0$) with consistent evidence. Evidence from Table 1 largely supports recency for series of short, complex evidence items (9 of 11 studies). However, evidence concerning SbS is sparse (only 2 studies) and, once again, it is not clear that any studies required evaluation ($R = 0$) with consistent evidence.

The third row of Table 2 concerns long series of evidence items for which the model predicts mixed results. On the one hand, SbS processing typically leads to recency. However, as more information is processed across time, we predict decrements in α and β that will eventually induce primacy. Thus, whereas recency may be observed in long series, this becomes less likely as the length of the series increases (for relevant experimental evidence, see Anderson, 1959). Interestingly, in one of the two studies reporting recency for long series of simple evidence items (Lichtenstein & Srull, 1987—see Table 1), subjects were specifically told that they would be asked to recall specific evidence items for comprehension. That this instruction disturbed the judgmental process by focussing attention on particular evidence items was borne out by the finding of primacy in the two other instructional conditions of this experiment (an impression set and a memory set). In the second case, whereas Jones, Rock, Shaver, Goethals, and Ward (1968) reported recency in one of six experiments (see Appendix A), replications by other researchers

(Feldman & Bernstein, 1977, 1978) resulted in primacy. Finally, Tetlock (1983) showed that, although primacy resulted for a long series of complex evidence items with the EoS response mode, order effects could be nullified by instructing subjects that they would be required to justify their decisions, i.e., by intervening in the process.

NEW EXPERIMENTAL EVIDENCE

Comparison of Tables 1 and 2 shows strong agreement between predictions and existing experimental findings. This is particularly the case for series of short, simple evidence items in row 1 of Table 2 for which there are many data (see Table 1). Of special interest are the predictions of row 2 of Table 2 where not only have fewer studies been conducted in the past (see Table 1), but where the present model makes novel predictions. These are the presence of recency under all conditions with the exception of the cells involving consistent evidence in evaluation tasks (where $R = 0$). We therefore focus our experimental testing of the model on the predictions implied by row 2 of Table 2.

Before presenting experimental evidence, it is appropriate to ask what alternative models could make the same predictions as our model. A "straw-man" alternative is the Bayesian model that has been much utilized to study belief updating (Slovic & Lichtenstein, 1971). However, since this predicts no order effects under any conditions, its descriptive adequacy is discredited by the studies summarized in Table 1 and Appendix A. A second class of models are those of Anderson (1959) and Carlson and Dulany (1988) which are identical in form to our SbS models when $R = S_{k-1}$ and $R = 0$, respectively. However, the former predicts recency for both mixed and consistent evidence, and the latter does not predict order effects for consistent evidence. Third, by changing assumptions concerning the adjustment parameter, w_k, it is possible to construct models that are variants of Eqs. (7a) and (7b) and which do make alternative predictions. For example, if w_k is either a constant or proportional to the scale value of evidence, $s(x_k)$, the model would predict no order effects for either mixed or consistent evidence. If, alternatively, w_k were proportional to $(1 - S_{k-1})$ for negative evidence and to S_{k-1} for positive evidence (thereby reversing our contrast assumption), the model would predict primacy for mixed evidence and no effect for consistent evidence. However, apart from the fact that all of these alternatives concerning w_k have awkward mathematical implications (S_k would no longer be bounded between 0 and 1), they do not appear to have been suggested in the literature. The fact remains that the belief-adjustment model both accounts for a wide range of circumstances in which order effects have been demonstrated and makes subtle predictions for cases which have not previously been considered. Moreover, these predictions are unique relative to alternative models that have been suggested in the literature as well as variants of the model itself.

Plan of Experiments

As noted above, our experimental testing focused on the predictions in row 2 of Table 2. In Experiments 1 through 4 we examined how new information changes beliefs in a task using complex evidence items under conditions involving both SbS and EoS response modes. The stimuli for these experiments involved the updating of causal hypotheses for which information was presented in a form that, we hypothesized, would require subjects to encode evidence as positive or negative vis-à-vis the hypothesis under consideration such that $R = 0$ and $-1 \leq s(x_k) \leq +1$. For these stimuli, the model predicts no order effects for consistent evidence and recency for mixed evidence. Experiments 1 and 2 were designed to test the prediction of no order effects for consistent evidence. Experiments 3 and 4 tested the prediction of recency effects for mixed evidence.

Whereas the model predicts no order effects for consistent evidence when $R = 0$, it does predict recency for consistent evidence when $R = S_{k-1}$, i.e., for estimation as opposed to evaluation tasks. This differential prediction was tested in Experiment 5. The information in one of the scenarios used in Experiments 1 through 4 was presented in two forms, one intended to engage subjects in evaluation and the other intended to engage subjects in estimation. We predicted no order effects for the former, but recency for the latter.

It is important to recall that the order-effect predictions are derived from qualitative assumptions concerning the *process* of belief updating. In evaluation tasks, for example, beliefs should be revised upward after seeing positive evidence and downward after seeing negative. In addition, the contrast assumption predicts that belief changes will be proportional (negative evidence) or inversely proportional (positive evidence) to the current anchor. Together, these process assumptions imply predictions about the patterns of data that should be exhibited and thus the manner by which order effects do or do not occur. A second objective of the experiments was to examine these additional, implied tests of the process predicted by the updating model.

EXPERIMENTS 1–3

Because of commonality of procedures, we describe Experiments 1–3 together. The main purpose of Experiments 1 and 2 was to test for order effects in the updating of beliefs based on *consistent* evidence in a *short* series of *complex* evidence items where the theory predicts no order effects for either SbS or EoS. Experiment 1 dealt with *positive* evidence, Experiment 2 with *negative* evidence. In Experiment 3 we tested the model's prediction that *mixed* evidence would lead to recency effects for both SbS and EoS.

Methods

Subjects. Subjects were recruited through advertisements placed around the University of Chicago. They were offered $5/hour for participating in an experiment on decision making. Their median age was 22.5 years and their mean educational level was 4.2 years beyond high school. There were 24 subjects in each of the three experiments.

Stimuli. The stimuli involved a set of four scenarios, each of which involved an initial description (the stem) and two additional pieces of information presented in separate paragraphs. Excluding response scales and instructions, the stems of the scenarios varied in length between 68 and 109 words (mean of 88) with additional pieces of information averaging 52 words each. The content of the four scenarios involved: (1) a defective stereo speaker thought to have a bad connection; (2) a baseball player whose hitting has improved dramatically after a new coaching program; (3) an increase in sales of a supermarket product following an advertising campaign; and (4) the contracting of lung cancer by a worker in a chemical factory. In each case, the stem provided information regarding the hypothesis that the particular cause was responsible for the effect of interest. After reading the stem, subjects were asked to rate how likely the suspected factor was the cause of the outcome on a rating scale from 0 to 100 (e.g., in the baseball scenario, subjects were asked, "How likely do you think that the new training program caused the improvement in Sandy's performance?"). After responding to this question, subjects turned the page of their experimental booklets and were presented with two pieces of additional information regarding the causal hypothesis.

In Experiment 1, these two pieces consisted of strong positive and weak positive information about the hypothesis. As an example, a positive argument in the baseball scenario read "The other players on Sandy's team did not show an unusual increase in their batting average over the last five weeks. In fact, the team's overall batting average for these five weeks was about the same as the average for the season thus far." To illustrate why we consider that this type of evidence would evoke evaluation as opposed to estimation, note, first, the hypothesis to be judged is dichotomous; i.e., did or did not the training program cause the improvement in Sandy's performance? Second, the potential impact of the evidence is bipolar and its sign (positive or negative vis-à-vis the hypothesis) must be interpreted prior to integration with the existing anchor. In addition, the evidence has not been neatly "packaged" in a quantitative, unipolar form that is compatible with estimation.

The new information was presented in either a strong–weak or weak–strong order. In the SbS condition, the two pieces of new information were presented on separate pages with a 0- to 100-point rating scale at the bottom of each page. Subjects were asked to respond after each piece to the question, "Now, how likely do you think X caused Y?" In the EoS procedure, the two pieces of information were presented continuously as paragraphs. At the end of the last paragraph, subjects were asked to rate the likelihood that X caused Y.

In Experiments 2 and 3, the same procedures were followed except that in Experiment 2 the two pieces of information consisted of strong negative and weak negative information about the hypothesis, and in Experiment 3 the two pieces of information were mixed involving positive and negative information. An example of negative information in the baseball scenario was "The games in which Sandy showed his improvement were played against the last place team in the league. Pitchers on that team are very weak and usually allow many hits and runs."

Design and procedure. The experimental design involved three variables: order of evidence (strong–weak vs weak–strong for Experiments 1 and 2, and positive–negative vs negative–positive in Experiment 3); response elicitation procedure, SbS vs EoS; and the four different scenarios. The first two variables were factorially crossed (resulting in four combinations) and arranged as within-subject variables. The four scenarios were presented in a 4×4 Latin square arrangement. Thus, subjects evaluated each of the four scenarios in one of the four combinations resulting from crossing the order and elicitation variables. The 24 subjects in each experiment were randomly assigned to one of the four groups making up the Latin square (6 subjects per group).

Subjects were given the experimental materials in booklet form and told to work carefully and at their own pace. To provide variety, they worked on another, unrelated experimental task between scenarios. On average, subjects completed all tasks in 1 h, working on the tasks in a laboratory under the supervision of an experimenter with at most three other subjects present at the same time. At the completion of the experiment, subjects were asked to reconsider each piece of evidence and rate it on a scale from -100 (completely disconfirms the hypothesis) to $+100$ (completely confirms the hypothesis). These data were used to provide a manipulation check on whether the information was of the hypothesized size and sign.

Results

We present three types of results. First, we discuss manipulation checks. Second, we consider order effects. Third, we go beyond order effects to examine the extent to which the patterns of data are consistent with the assumed underlying process of the model and, in particular, the contrast assumption.

Manipulation checks. In Experiment 1, across all four scenarios, the mean rating of the strong positive evidence was 63 while the weak positive evidence was rated as 33, $t(92) = 9.0, p < .001$. Furthermore, in each scenario the strong evidence was rated higher than the weak evidence, $t(23) > 3.5, p < .001$, for all scenarios. In Experiment 2, over the four scenarios, strong negative evidence was rated as being more negative than weak negative evidence, -30 vs -12, $t(92) = -5.14, p < .001$. This result also held in all four scenarios, although at different levels of statistical significance: for two scenarios, $t(23) > 3.5, p < .001$; for the third, $t(23) = 1.7, p < .10$; and for the fourth, $t(23) = 1.4, p < .17$. In Experiment 3, over all four scenarios, the positive evidence was rated 65, $t(92) = 24.0, p < .001$, and the negative evidence was rated -38, $t(92) = -11.2, p < .001$. In addition, this pattern held in each of the four scenarios, $t(23) > 3.5, p < .001$.

As an additional test, consider how many times subjects revised opinions down after seeing positive evidence and up after receiving negative evidence during the experiments. Across all three experiments there were 285 judgments in the SbS condition that should have been constrained in sign (288 less 3 missing observations). Nineteen of these 285 judgments or 6.7% were in the "wrong" direction. In the EoS conditions of Experiments 1 and 2, 7 of 96 or 7.3% of final judgments involved errors in direction (up after consistent negative evidence or down after positive).

Further evidence of subjects' sensitivity to the relative strength of the weak and strong information presented in Experiments 1 and 2 can be noted by whether belief changes following strong evidence were greater than after weak. Because of the contrast assumption (see Eqs. (6a) and (6b)), the model implies that this should always be the case for the strong–weak condition but not necessarily for the weak–strong and was verified in 35 of 47 or 75% of individual cases that could be tested in Experiments 1 and 2 for the strong–weak order, $z = 3.21, p < .001$, one-tailed sign test. The analogous statistics in the weak–strong order were 30 of 47 or 64%, $z = 1.75, p < .05$, one-tailed sign test. Taken together, the above three sets of results indicate that we were successful in manipulating both the sign and the differential strength of the evidence in all three experiments.

Order effects. Figures 2, 3, and 4 illustrate graphically the results concerning order effects in Experiments 1, 2, and 3, respectively, for both the SbS and EoS conditions. As predicted, the figures are consistent with no order effects in Experiments 1 and 2 but recency in Experiment 3.

One difficulty in analyzing order-effect data when treatments concerning order are between subjects is that initial differences can mask effects (consider, for example, Fig. 3b). For statistical analysis, therefore, we focus on the difference between the stem (S_0) and final judgments (S_2), i.e., $Y = S_0 - S_2$. In each experiment, this measure was subjected to a $2 \times 2 \times 4$ analysis of variance using the appropriate repeated-measures Latin square design. In Experiment 1, only one effect was significant, a main effect for scenarios, $F(3,72) = 12.9, p < .001$. This was due to the fact that one scenario (the stereo speaker) increased more than the others ($Y = -31$ vs -9, -10, and -12). From our perspective, the major

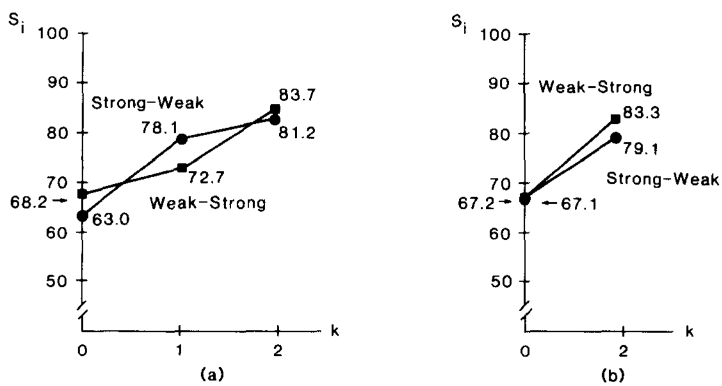

FIG. 2. Opinion curves of Experiment 1: (a) Step-by-Step response mode. (b) End-of-Sequence response mode.

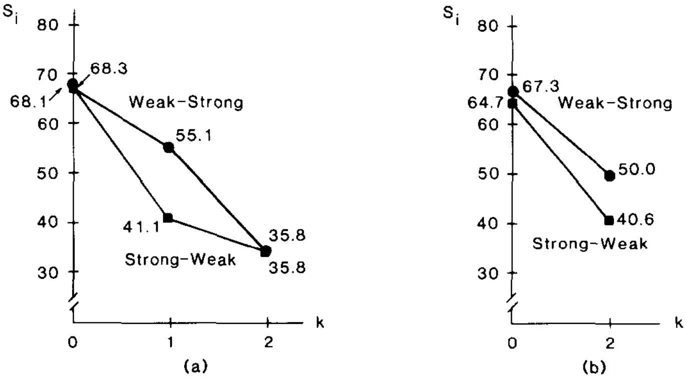

FIG. 3. Opinion curves of Experiment 2: (a) Step-by-Step response mode. (b) End-of-Sequence response mode.

finding was the predicted lack of an order effect, the mean increases for the strong–weak and weak–strong orders being 15.2 and 15.8, respectively (Figs. 2a and 2b).

In Experiment 2, the $2 \times 2 \times 4$ analysis of variance on $S_0 - S_2$ also showed no effect of order, in accord with our prediction. Mean decreases for strong–weak and weak–strong orders were 27.8 and 28.8, respectively. However, there was a main effect for response mode, $F(1,70) = 3.9$, $p < .05$, and a response mode x scenario interaction, $F(3,70) = 3.5$, $p < .02$. The main effect occurred because the initial judgments decreased by 32 under the SbS procedure versus 25 under the EoS method. The interaction was due to the fact that one scenario (the stereo speaker) had a larger decrease in the EoS method than the three other scenarios (which showed larger decreases in the SbS procedure). Since we did not control

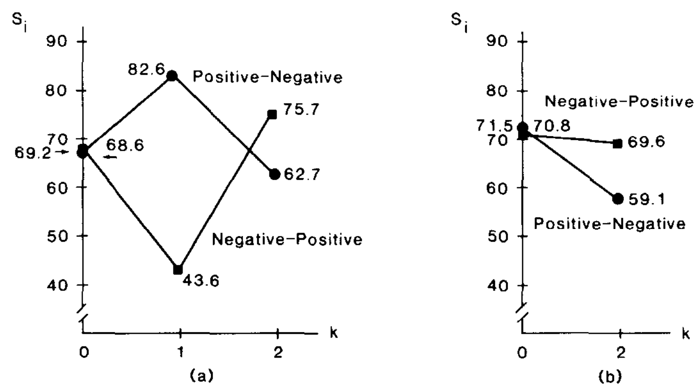

FIG. 4. Opinion curves of Experiment 3: (a) Step-by-Step response mode. (b) End-of-Sequence response mode.

for the variables constituting scenario content (indeed, it is not even clear what these variables are), the differential sensitivity of belief change to particular scenarios under various procedures raises interesting questions that are beyond the scope of the present model.

In Experiment 3, the $2 \times 2 \times 4$ analysis of variance showed the hypothesized recency effect, $F(1,71) = 9.8$, $p < .003$. Specifically, the positive–negative order resulted in a mean decrease of 9.2 relative to the initial judgment; the negative–positive order resulted in an increase of 2.7. In addition to the main effect for order, there was a significant scenario main effect, $F(3,71) = 9.8$, $p < .001$, and a scenario x response mode interaction, $F(3,71) = 2.7$, $p < .05$, that were similar to those found in Experiments 1 and 2.

The absence or existence of order effects at the aggregate level does not necessarily imply the same outcomes at the level of individual subjects. However, because in each experiment each subject saw four different scenarios, it is not possible to make direct tests of order effects at the individual level. On the other hand, within both the SbS and the EoS conditions, each subject was exposed to different orders of evidence, namely strong–weak vs weak–strong in Experiments 1 and 2, and positive–negative vs negative–positive in Experiment 3. Thus, by ignoring scenario differences a *surrogate* test of order effects can be made at the individual subject level. To do so, responses of individual subjects were classified within both the SbS and the EoS conditions to see whether they indicated primacy, recency, or no order effects (absolute sizes of effects were ignored). Results are summarized in Table 3. For Experiments 1 and 2, there is an almost exact 50/50 split between primacy and recency for both the EoS and the SbS conditions. The null hypothesis of no order effects therefore cannot be rejected for Experiments 1 and 2, $z = 0$ and $z = .147$, ns by sign tests. For Experiment 3, on the other hand, the same test reveals a majority of recency effects in both the SbS (61%) and the EoS (75%) conditions, $z = 2.51$, $p < .01$, one-tailed sign test. Although

TABLE 3
Surrogate Individual-Subject Level Order Effects

	Experiment 1			Experiment 2			Experiment 3		
	SbS	EoS	Total	SbS	EoS	Total	SbS	EoS	Total
Primacy	11	12	23	12	11	23	8	6	14
Recency	10	12	22	11	12	23	14	18	32
No effect	1	—	1	1	1	2	1	—	1
Missing	2	—	2	—	—	—	1	—	1

Note. Table entries are numbers of subjects showing different "surrogate" order effects at the individual level.

not ideal, these tests of individual-level order effects add further support to the model's experimental predictions.

Tests of the contrast assumption. A key feature of the underlying model is the contrast assumption: Negative evidence induces greater belief change the larger the preceding anchor; for positive evidence it is the reverse—see Eqs. (6a) and (6b). The contrast assumption therefore imposes restrictions on patterns of data that are consistent with the model:

(i) For consistent evidence, an item of evidence will induce more change in belief the earlier it is processed. To see this, consider the effects of two pieces of negative evidence. The piece that is considered second will necessarily adjust a lower anchor than that processed first. Thus, had the second piece been processed first, it would have adjusted a larger anchor thereby inducing greater change in belief. In Experiments 1 and 2, this implies that strong evidence should induce more change in belief in the strong–weak than in the weak–strong order. Similarly for weak evidence, greater changes in belief should be observed in the weak–strong rather than the strong–weak order.

(ii) For mixed evidence, evidence will induce more belief change when processed after evidence that is opposite in sign. This follows from the fact that positive evidence processed in a negative–positive order has a lower anchor than when processed in a positive–negative order. Similarly, negative evidence processed in a positive–negative order has a larger anchor than when processed in a negative–positive order.

(iii) The amount of belief change, defined by $(S_{k-1} - S_k)$, should be proportional to the preceding anchor, S_{k-1}—see Eqs. (6a) and (6b). This should be evidenced by positive correlation between $(S_{k-1} - S_k)$ and S_{k-1}. (To see this, note that the amount of belief change associated with positive evidence should be inversely proportional to S_{k-1}; in addition, $(S_{k-1} - S_k) < 0$. For negative evidence, belief change should be proportional to S_{k-1}, but $(S_{k-1} - S_k) > 0$.)

We now consider tests of implications (i) and (ii) at the aggregate level. For positive evidence, examine Fig. 2a and note that when the strong evidence is presented first (in the strong–weak order) it increases mean opinion by 15.1, from 63.0 to 78.1. When it appears second (in the weak–strong order) it induces a smaller increase of 11.0, 72.7 to 83.7. Similarly, the weak evidence has greater impact with the lower anchor in the weak–strong as opposed to the strong–weak order, i.e., $(72.7 - 68.2) > (81.2 - 78.1)$. Although in the appropriate direction, however, post hoc contrasts show that neither of the differences in impacts are statistically significant, $t(44) = .86$ and $t(44) = .48$, respectively.

Figure 3a illustrates the data from the SbS condition of Experiment 2. The strong negative evidence discounts belief more from the larger anchor in the strong–weak as opposed to the weak–strong situation, i.e., (68.1 − 41.1) > (55.1 − 35.8), $t(42) = 1.58$, $p = .06$, one-tailed test. Similarly, the weak evidence has greater impact from the larger anchor, i.e., (68.3 − 55.1) > (41.1 − 35.8), $t(42) = 2.16$, $p < .025$, one-tailed test.

Figure 4a shows the analogous data for Experiment 3 with respect to mixed evidence. As predicted, the increase in mean opinion due to positive evidence is larger from the lower anchor, i.e., (75.7 − 43.6) > (82.6 − 68.6), $t(43) = 3.07$, $p < .025$. However, the implication is not validated for negative evidence. Here, the negative evidence has more impact when discounting a smaller anchor, i.e., (69.2 − 43.6) > (82.6 − 62.7) but the difference is not statistically significant, $t(43) = -1.09$.

In summary, the predicted directions of five of the six tests of implications (i) and (ii) (across the three experiments) are validated by the aggregate data. The exception concerns the use of negative evidence in Experiment 3 where the model would have predicted more discounting of the negative evidence in the positive–negative as opposed to negative–positive order. The failure to observe more discounting in this instance can be interpreted as meaning that the recency effect exhibited in Fig. 4a is smaller than would have been predicted by the model.

Whereas it is not possible to test predictions (i) and (ii) directly at the individual level, the recency effect in Experiment 3 can be examined further by asking whether the change in belief due to processing the second piece of evidence exceeds that due to the first.[4] The individual data support this implication in the SbS condition. The change in belief due to the second piece of evidence exceeded that due to the first in 32 of 47 cases, 68%, $z = 2.33$, $p < .01$, one-tailed sign test.

Finally, tests of implication (iii) are provided in Table 4 which shows correlations between changes in belief, $(S_{k-1} - S_k)$, and their preceding anchors, S_{k-1}. In addition to the data from the SbS conditions, we also include statistics from the EoS data where the correlation is necessarily calculated between $(S_0 - S_2)$ and S_0. In order to have meaningful sample sizes, the correlations have been calculated ignoring scenario differences. All 16 correlations are of the appropriate sign and 11 of the 16 are statistically significant.

In summary, the data from the three experiments conform with the order-effect predictions of the model, both in the aggregate and in our "surrogate" tests of individual-level effects. As to the process inducing order effects, both direct and indirect tests showed that subjects interpreted and acted on the sign and strength of the evidence in the manner predicted by the model. In addition, most tests of the implications of the contrast assumption were confirmed.

[4] This is a weaker test because it assumes that $s(+) = s(-)$ and $\alpha = \beta$.

TABLE 4
Experiments 1–3: Correlations between S_{k-1} and $(S_{k-1} - S_k)$

	Experiment 1		Experiment 2		Experiment 3	
	S/W	W/S	S/W	W/S	P/N	N/P
			Step-by-Step			
Correlation between						
S_0 and $(S_0 - S_1)$	.679**	.512**	.391*	.230	.554**	.222
S_1 and $(S_1 - S_2)$	.622**	.644**	.251	.431*	.247	.644**
			End-of-Sequence			
Correlation between						
S_0 and $(S_0 - S_2)$	.663**	.636**	.328*	.209	NA	NA

Note. S/W = strong–weak order; W/S = weak–strong order; P/N = positive–negative order; N/P = negative–positive order.
** $p < .005$.
* $p < .050$.

EXPERIMENT 4

Experiment 4 was designed to test the prediction of recency for mixed evidence but involved four as opposed to two pieces of evidence. Parts of the various scenarios used in Experiments 1–3 were, however, expressed differently.[5]

Methods

Subjects. There were 60 subjects from the same population as the previous experiments. They were recruited and remunerated in the same manner.

Stimuli, design, and procedures. Experiment 4 followed Experiment 3 in all respects except that there were four rather than two pieces of evidence and, as noted above, parts of the scenarios were longer as measured by number of words. The two orders were therefore (+, +, −, −) and (−, −, +, +). Within the two positive and negative pieces, the orders were held constant.

Results

The manipulation checks showed that all positive arguments were seen as positive and all negative arguments seen as negative. All means were significantly different from zero and in the hypothesized directions, $t(59) > 3.5, p < .001$. The individual data were also checked to note the number of times beliefs changed in a direction opposite that intended by us (i.e., down after seeing positive evidence or up after negative). Of the 479 judgments that should have been constrained in sign (480 minus one missing observation), there were 48 such violations, a rate of 10%.

[5] Experiment 4 was actually conducted before Experiments 1–3. The differences in wording had been made to clarify potential ambiguities of interpretation in some of the stimuli.

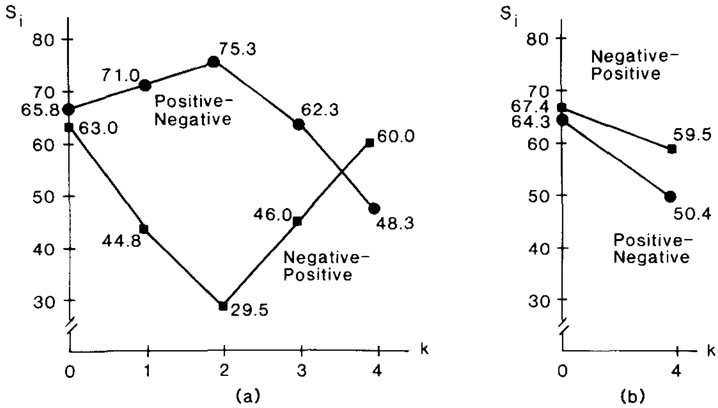

FIG. 5. Opinion curves of Experiment 4: (a) Step-by-Step response mode. (b) End-of-Sequence response mode.

Figures 5a and 5b depict the pattern of the aggregate data and reveal the predicted recency effects for both the SbS and EoS conditions as measured by differences in $S_0 - S_4$, means of 14.5 and 6.0, respectively. Using $Y = S_0 - S_4$ as the dependent variable, analysis of variance showed that the overall order effect was statistically significant, $F(1,180) = 17.48$, $p < .001$. As in the previous experiments, there was a significant scenario effect, $F(3,180) = 7.11$, $p < .001$, as well as a scenario by response mode interaction, $F(3,180) = 4.90$, $p < .003$.

Because of scenario differences, it is not possible to make direct tests of order effects at the individual level. However, surrogate individual-level tests were made in the same manner as the previous experiments. In the SbS condition, 42 of 57 or 74% of individual tests showed recency as opposed to primacy (there were 2 with no effect and 1 missing observation). In the EoS condition, 36 of 57 or 63% showed recency as opposed to primacy (there were 3 with no effect). One-tailed sign tests showed surrogate individual-level recency effects to be statistically significant for both response modes, $z = 3.46$, $p < .001$, and $z = 1.86$, $p < .05$, respectively.

Concerning the overall pattern of the aggregate data in the SbS condition, the contrast assumption implies predictions concerning the relative sizes of changes in belief. In particular, the increase due to positive evidence should be larger in the negative–positive order than in the positive–negative order because the anchor associated with the latter is larger than that for the former. Indicating anchors in the negative–positive order by primes and those in the positive–negative order without primes, this prediction can be written: (i) $(S_1 - S_0) < (S_3' - S_2')$—see Fig. 5a. Using analogous arguments, the following predictions can also be made: (ii) $(S_2$

$- S_1) < (S_4' - S_3')$; (iii) $(S_2 - S_3) > (S_0' - S_1')$; and (iv) $(S_3 - S_4) > (S_1' - S_2')$—see Fig. 5a. Means of these differences are, respectively, (i) 5.00 vs 16.58, $z = -3.65, p < .001$; (ii) 4.33 vs 13.51, $z = -2.95, p < .025$; (iii) 13.00 vs 18.12, $z = -1.54$, ns; and (iv) 13.92 vs 15.51, $z = -0.55$, ns. Thus, (i) and (ii) are validated and statistically significant but the data do not support (iii) and (iv). The substantive implication of these tests is that although the aggregate data show the predicted recency effect, this is smaller than would have occurred had the predicted relative sizes in changes in belief (i) through (iv) all been validated.

Further tests of the model are implied by expected correlations between anchors and subsequent changes in belief. Specifically, $(S_{k-1} - S_k)$ should be positively correlated with S_{k-1} in the SbS data. For $k = 1$ to 4 in the positive–negative condition, these correlations were .384, .473, .275, and .441, respectively, $p < .025$, for all. In the negative–positive condition, the analogous statistics were .277, .580, .262, and .152, $p < .025$, for all except the last correlation which was not statistically significant. In other words, all correlations were of the appropriate sign, and seven of eight were statistically significant.

EXPERIMENT 5

Experiments 1 and 2 tested the prediction that, for evaluation tasks (i.e., where $R = 0$), no order effects would be observed for consistent evidence. These results are subject to alternative explanations. First, if order effects do occur for consistent evidence, they are likely to be smaller than those for mixed evidence. This could occur because differences between $s(x_k)$ values for consistent evidence would tend to be smaller than those between mixed evidence. Second, our tests might have lacked statistical power. In addition, whereas no order effects are predicted for consistent evidence in evaluation tasks, the model does predict recency for estimation tasks where $R = S_{k-1}$ (see Table 2). Moreover, there is evidence in the literature that order effects have been observed with consistent evidence (Anderson, 1968; Chalmers, 1971).

Experiment 5 examined these issues. We designed an updating task in which consistent evidence was presented in alternative forms. For one version, we hypothesized subjects would use an evaluation strategy and that this would result in no order effects thereby replicating the outcomes of Experiments 1 and 2. For the second version, we hypothesized that subjects would use an estimation strategy such that there would be an order effect, specifically recency (see Table 2).

To interpret both the motivation for this experiment and our subsequent results, it is important to emphasize a qualitative difference in the way consistent evidence is processed in the evaluation and estimation modes. Imagine a task requiring step-by-step responses for two pieces of positive evidence, one weak and one strong. When evidence is processed in the weak–strong order, the evaluation and estimation modes imply upward adjustments after processing both the weak and the strong evidence. However, when the evidence is processed in the strong–weak order this is no longer the case. On the one hand, both modes do imply upward adjustments after the strong evidence. On the other hand, whereas the weak evidence implies an upward adjustment in the evaluation mode, its effect is a downward adjustment in the estimation mode. This occurs because whereas the sign of the adjustment is determined relative to $R = 0$ in the evaluation mode, $R = S_{k-1}$ in the estimation mode. Thus, if weak evidence follows strong in the estimation mode, it will be seen as weaker than S_{k-1} thereby implying a downward adjustment. We therefore further predict that the directions of the adjustment following receipt of the weak evidence in the strong–weak order will be positive in the evaluation task but negative in the estimation task.

Methods

Subjects. There were 60 subjects who were recruited through advertisements placed around the University of Chicago. They were paid $2 for participating in this experiment after having completed another unrelated experiment. Their mean age was 21 years.

Stimuli. The stimuli were adapted from the defective stereo scenario used in the preceding experiments. In this scenario, subjects are asked to assess the likelihood that the malfunction in the protagonist's (Judy's) stereo is due to a loose connection between the speaker and the amplifier. In our earlier discussion, we hypothesized that use of evaluation as opposed to estimation depends on (a) the nature of the opinion to be expressed (dichotomous belief vs estimates on a continuous scale) and (b) the extent to which the evidence presented is compatible with the two forms of encoding. We specifically argued that estimation is facilitated when evidence is presented on an implicit, unipolar scale that maps easily onto the required judgment, whereas evaluation occurs when the potential impact of the evidence is bipolar and its sign (positive or negative vis-à-vis the hypothesis) must be interpreted prior to integration with the existing anchor.

To operationalize conditions that would invoke different modes of encoding, we constructed two versions of the same scenario. Whereas both versions required responses concerning the same dichotomous opinion, the evidence was presented in different forms. For the estimation mode, we characterized the evidence using explicit numerical information on a unipolar scale that was compatible with the response mode. For the evaluation mode, we used verbal information that required explicit interpretation concerning the sign of the evidence.

For the weak evidence, the information in the estimation [evaluation] version was "Judy has a friend who knows a lot about stereo systems. She told him about her problem. He said it was quite common and that 60% of the time [more often than not] the cause turned out to be a loose connection between the speaker and the amplifier." That is, the only difference between the two versions lay in the words "60% of the time" versus "more often than not." The information for the strong evidence in the estimation [evaluation] version was "Judy asked a salesman in a stereo store about her problem. He told her that others had asked him about the same problem and that 80% of the time [in the vast majority of cases] the cause turned out to be a loose connection between the speaker and amplifier."

Whereas the differences between the evaluation (or verbal) and estimation (or numerical) versions of the stimuli were small, we hypothesized that they would evoke different processes. For the evaluation task, subjects must read the evidence and evaluate whether it is positive or negative vis-à-vis the hypothesis under consideration before making their judgments. As observed in the previous experiments, this typically leads to an adjustment in the direction indicated by the evidence. Moreover, this should lead to no order effects with consistent evidence. In the estimation task, however, the evidence is already encoded in an easy-to-use, unipolar, quantitative form that is compatible with estimation. Even with consistent evidence, this should lead to recency effects.

Design and procedure. The experimental design involved two between-subjects variables each with two levels: order of evidence (strong–weak vs weak–strong) and type of information (numerical vs verbal). Fifteen subjects were allocated at random to each of the four cells of the design. Subjects were given the experimental materials in booklet form and instructed to work carefully and at their own pace but without flipping back to previous answers. The stem and each piece of positive evidence were on succeeding pages in the booklet and subjects made responses in step-by-step fashion on each of these three pages on a scale from 0 ("A loose connection does not cause the malfunction") to 100 ("A loose connection causes the malfunction"). After completing these judgments, subjects were asked to reconsider each piece of evidence and rate it on scale from −10 ("Totally undermines the hypothesis") to +10 ("Provides irrefutable positive evidence").

Results

We first consider manipulation checks based on the judgments made after completing the main task. For both the numerical and the verbal stimuli, subjects judged the strong evidence as stronger than the weak evidence, means of 5.02 vs 3.54 for the numerical, $t(29) = 3.13, p < .01$, and means of 5.38 vs 4.38 for the verbal, $t(29) = 2.51$, $p < .02$. In addition, judgments of the strong evidence did not differ between the numerical and the verbal versions of the stimuli, $t(58) = -0.57, p = .57$, and the same result held for the weak evidence, $t(58) = -1.19, p = .24$.

The pattern of experimental results is exhibited in Figs. 6a and 6b and conforms with predictions. There is a recency effect for the numerical stimuli but no order effect for the verbal stimuli. For the numerical stim-

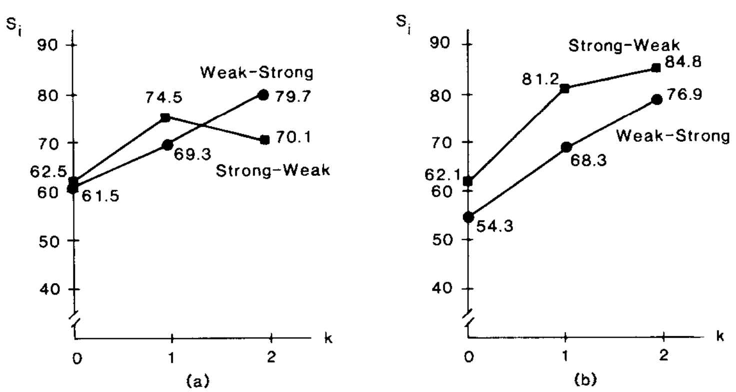

FIG. 6. Opinion curves of Experiment 5: (a) Numerical stimuli. (b) Verbal stimuli.

uli, the mean difference between the initial and the final judgments is 18.1 in the weak–strong order vs 7.7 in the strong–weak order, $t(28) = 2.01$, $p = .027$, one-tailed test. The analogous differences for the verbal stimuli are 22.6 and 22.7, $t(28) = .013$, $p = .990$. In addition, note from Fig. 6 that whereas the final judgment in the strong–weak order results in a drop from the preceding judgment for the numerical stimuli, there is an increase in the final judgment for the verbal stimuli. Both the mean reduction of -4.33 (numerical stimuli) and increase of 3.60 (verbal stimuli) are significantly different from 0, $t(14) = -3.04$, $p < .01$, and $t(14) = 3.58$, $p < .01$, respectively.

The aggregate pattern of results is supported by individual-level analysis. As in the previous experiments, the theory predicts the direction of changes in judgment following the receipt of new evidence. Overall, there were 6 violations of a possible 120 (i.e., where adjustments are in the direction opposite to that predicted). Of particular significance is the manner in which subjects in the strong–weak order made their final adjustments, i.e., for the weak evidence. Of the 15 subjects in the numerical condition, none adjusted upward. In the verbal condition, only 1 of 15 made a downward adjustment.

DISCUSSION AND FURTHER EVIDENCE

Experiments 1–4 tested predictions for task conditions for which little prior work has been reported in the belief-updating literature (see Tables 1 and 2). The results are consistent with the predictions of the belief-adjustment model concerning order effects for short series of complex evidence items when $R = 0$, vis., no order effects in the processing of consistent evidence (positive and negative), but recency for mixed evidence. Moreover, and also as predicted, these results hold across both the SbS and the EoS response modes. In addition to the order-effect predictions, implications of the contrast assumption were tested and generally found to hold.

Experiment 5 was a test of differential predictions concerning order effects for consistent evidence in evaluation and estimation. As predicted, consistent evidence lead to recency in estimation (the numerical stimuli) and to no order effects in evaluation (the verbal stimuli). Detailed analysis of the data further distinguished the implications of the two modes of encoding in respect of evidence processed in the strong–weak order. As predicted, the direction of belief revision following the weak evidence differed between modes, up for evaluation but down for estimation. Moreover, these findings indicate that the experimental methods used were powerful enough to detect order effects in consistent evidence which were absent in Experiments 1 and 2.

Our model has also been the subject of independent investigation by other researchers. In a series of four experiments, Ashton and Ashton (1988) tested for the absence of order effects for consistent evidence (positive and negative) and the presence of recency with mixed evidence. They used an SbS response mode and each experiment involved four pieces of evidence. Their subjects were advanced students in auditing and professional accountants and were required to perform updating tasks in an auditing content. Results were as predicted by the belief-adjustment model (i.e., no order effects for consistent evidence, recency for mixed).

In addition to order-effect predictions, Ashton and Ashton (1988) tested implications of our contrast assumption concerning the adjustment parameter, w_k, in two experiments in which they provided three groups of subjects with different anchor points, .20, .50, and .80, and subsequently presented the same evidence to all subjects. In one experiment, subjects received four items of positive evidence, in the other four pieces of negative evidence. Changes in opinion followed the predicted pattern with the amount of belief change being inversely proportional to the size of the anchor for positive evidence, but proportional for negative evidence. Specifically, in the positive condition the amounts of belief change were .266, .207, and .043 for the .20, .50, and .80 cases, respectively. In the negative condition, the corresponding amounts of belief change were $-.077$, $-.288$, and $-.361$. One could argue that, for the .80 anchor in the positive condition and the .20 anchor in the negative condition, the amounts of belief change were constrained by the boundedness of the 0–1 scale. However, the predicted differences were also observed between the .20 and the .50 anchors conditions for positive evidence and the .50 and the .80 for negative evidence, where scale boundedness should not be an important variable. A further implication tested by Ashton and Ashton was that of the relation between the magnitude of recency effects and the size of the initial anchor. The model predicts no relation (see Appendix C) and none was found.

A further study of updating by professional accountants in an auditing task was conducted by Tubbs, Messier, and Knechel (1990). They investigated both SbS and EoS response modes in tasks using consistent positive, consistent negative, and either two or four pieces of mixed evidence. The model predicts no order effects with consistent evidence and none was observed except for a small recency effect for positive information in SbS. With mixed evidence, the model predicts recency, and this was observed in three cases, the exception being an absence of order effect for EoS with two pieces of evidence. Subsequent studies using professional auditors as subjects have further validated the prediction of recency for short series of complex evidence items involving mixed evidence (Asare, 1989; Messier, 1989; Messier & Tubbs, 1990; Koch, Pei, & Reed, 1989). An interesting feature of the Koch, Pei, and Reed study was an attempt to measure whether the subjective evaluation of evidence, i.e., $s(x_k)$, was affected by serial position. In accordance with the assumptions of the belief-updating model, no such effects were found. Finally, in a study by Butt and Campbell (1989), professional auditors exhibited only small recency effects which also interacted with whether they had high or low priors (i.e., initial anchors) concerning the hypothesis. Interestingly, this study involved updating 10 pieces of evidence so it approaches what we would call a long series of items. Recall from Table 2 that as the number of items increases, recency effects are predicted to diminish.

In a study involving undergraduate business students, Dillard, Kauffman, and Spires (in press) both replicated the order-effect predictions we obtained in Experiments 1–4 using our stimuli and reported further support for both the contrast assumption and the implication of independence between level of the initial anchor or prior opinion and the size of recency effects. Dillard et al. also obtained recency effects for mixed evidence using management accounting scenarios. A further validation of recency effects using our stimuli for an SbS response mode was reported by De Wit, Hogarth, Koehler, and Luchins (1989).

Our model can also explain other phenomena of belief revision. In a series of ingenious experiments, Lepper, Ross, and colleagues (for an overview, see Ross & Lepper, 1980) have demonstrated a phenomenon of belief perseverance. In this paradigm, subjects are induced into acquiring certain beliefs, the source of which is subsequently discredited. However, instead of beliefs reverting to their initial levels, these investigators find that people persevere in their newly acquired but discredited beliefs. This phenomenon can be accounted for within our framework by assuming that $\beta > \alpha$ (i.e., people become advocates for the positions with which they have been endowed) but does not, of course, explain why this occurs. In a further test of this phenomenon in a mock trial situation, however, Hatvany and Strack (1980) found a different effect. Subjects changed their beliefs about a case on hearing a witness's testimony. But, when that testimony was subsequently discredited, subjects readjusted their beliefs to such an extent that the net effect of the witness's testimony was negative.

The Hatvany and Strack (1980) finding can be explained parsimoniously within the structure of the SbS process model when $R = 0$, and even when $\alpha = \beta$ (i.e., when people have the same attitudes toward positive and negative evidence). Imagine a case in which a person has an initial belief, S_0, and receives a piece of positive evidence which is subsequently discredited. Assume further that in discrediting this information, its subjective evaluation, denoted $s(x)$, does not change. However, because of the contrast assumption concerning the adjustment parameter, w_k, the amount by which opinion rises and then falls will not necessarily be the same. The reason is that each rise and fall in opinion is a function not only of the evidence, $s(x)$, but also of the previous anchor. Thus, whereas the amount by which the positive evidence increases S_0 will be equal to $\beta(1 - S_0)s(x)$ (see Eq. (7b)), the amount by which the evidence will lower the new anchor (S_1) is $\alpha S_1 s(x)$ (see Eq. (7a)) which, because $S_0 \neq S_1$, cannot be the same except when $S_1 = 1 - S_0$. This implies, for example, that if S_0 is high, the upward adjustment from incorporating the evidence will be more than compensated for by the downward revision when it is subsequently discredited so that the final opinion will be lower than S_0. Similarly, if S_0 is initially low, incorporation of positive evidence followed by discrediting the same will lead to a final opinion that is greater than S_0.

We call this the *asymmetric rebound effect* to denote that discrediting information does not necessarily imply returning to one's original position, and we emphasize that, within the model, this can occur even if there is no asymmetry in a person's attitude toward negative and positive evidence (i.e., $\alpha = \beta$).

GENERAL DISCUSSION

We now discuss our work in relation to the structure of updating models, limitations of our approach and opportunities for future research, and the importance of developing a procedural theory of judgment.

Structure of Updating Models

The basic structure of belief updating proposed in this paper is an anchoring-and-adjustment process that has the advantage of being relatively easy to implement from a cognitive viewpoint. Three critical features of our analysis center on assumptions concerning (1) the encoding of evidence, (2) the mode of processing, and (3) the adjustment weight.

(*1*) *Encoding*. In our framework, evidence is encoded relative to a reference point or background, R, which is equal either to the prior anchor, S_{k-1}, or a constant. Which of these reference points is adopted depends on whether people treat the task as involving evaluation or estimation. In *evaluation*, where evidence is thought of as bipolar in nature (e.g., supports or refutes a hypothesis), data are encoded as positive or negative with respect to the hypothesis under consideration. This was operationalized by allowing the subjective evaluation of evidence, $s(x_k)$, to vary between -1 and $+1$ and by setting $R = 0$. A consequence of this formulation is data that fit adding models of judgment. By contrast, in *estimation* sequential integration of information across time involves a "moving average" that is adjusted by new evidence in the direction in which this deviates from the present position. In these cases evidence is thought of as unipolar such that $s(x_k)$ varies between 0 and 1 and $R = S_{k-1}$ (the previous anchor). A consequence of this formulation is a model that produces averaging data of judgments (see also Lopes, 1982, 1987a).

As demonstrated above (Table 2 and our experimental work), whether people adopt an evaluation or estimation mode affects the kinds of order effects observed. Moreover, an important contribution of the evaluation/estimation distinction is to resolve the apparent conflict between adding or averaging formulations of judgmental processes. Consider, for example, what Anderson (1981) refers to as the "set size effect." Imagine a situation where a person is presented with, say, three pieces of information (in SbS format) each having the same evidential impact, i.e., of the same strength and polarity. What is the final opinion after seeing these pieces of evidence? According to an averaging model, the final opinion must be equal to the average of the scale positions of the evidence, which in this case is the same as the scale value of each of the three pieces of evidence. An adding model, however, predicts a final opinion that grows more extreme (i.e., higher or lower according to circumstances) than this common value. And indeed, this is what is observed empirically. To account for this phenomenon within an averaging formulation, Anderson has suggested that, in addition to the evidence items presented, subjects had in mind a previous value on the opinion scale that was more extreme than the evidence presented. Thus, when this imaginary prior value is averaged with the scale values of the information presented, the new average becomes more extreme across time thereby producing the so-called "set size effect."

Our model suggests two alternative explanations. One is that evidence is being encoded relative to a constant as opposed to varying reference point (i.e., $R = 0$ rather than $R = S_{k-1}$). The second is to note that people can approach judgmental tasks in ways that produce both "averaging" and "adding" data. Moreover, as demonstrated in Experiment 5, the differences in task conditions that invoke estimation or evaluation modes can be quite subtle, e.g., numerical versus verbal representations of the same information. Thus, although it would be convenient from a conceptual viewpoint to assume use of a single mode within the same task, the fact that apparently small contextual differences can evoke different processes makes plausible the notion that subjects might also shift modes within the same task. Thus, when subjects see evidence with the same scale value on successive occasions, they may switch from estimation to evaluation mode thereby avoiding having to repeat the same answer. Indeed, Lopes (1987a) has made a similar suggestion.

To summarize, an important contribution of the belief-adjustment model is to locate the substantive, process difference between models of judgment that produce averaging and adding data in the manner in which evidence items are encoded prior to integration in judgment (cf., Lopes, 1982). Moreover, because people use both evaluation and estimation modes of encoding, it may be unrealistic to believe that they only use one or the other within the same task. In the present work, we made specific hypotheses about conditions that tend to favor the use of each mode and provided an experimental demonstration of the difference between conditions. However, more work is required to provide complete specification of these conditions.

(2) *Processing mode.* Our model distinguished between two modes of processing, Step-by-Step and End-of-Sequence. Moreover, we examined tasks where subjects were required to respond either in a step-by-step basis or only after all the information had been received. In the presence of a step-by-step response mode, it is reasonable to assume that people must use an SbS process. However, an end-of-sequence response mode may or may not invoke an EoS process. By appealing to the well-established notion of cognitive limitations, we assumed that people will use an EoS process in the presence of an end-of-sequence response mode if this does not exceed their processing abilities. With more complex and longer tasks, however, they will adopt an SbS process. These assumptions raise interesting and, as yet, unanswered questions concerning the point at which people decide to switch from one strategy to another (i.e., EoS to SbS) as well as details concerning the EoS process that were not addressed in this paper. For example, what rules govern the aggregation of information that is used to adjust the anchor when using the EoS process? (See Eqs. (5) and (8), also Anderson, 1981.)

A further issue with respect to processing concerns the evaluation of evidence. In the normative Bayesian model of inference, evidence is evaluated with respect to both the hypothesis under consideration and its alternatives. In constructing our model, however, we explicitly built on the notion that people typically fail to evaluate evidence with respect to its diagnostic impact on alternative hypotheses. Rather, they tend to evaluate evidence with respect to a single hypothesis which is cognitively simpler (see, e.g., Schum & Martin, 1980; Fischhoff & Beyth-Marom, 1983; Robinson & Hastie, 1985). An intriguing issue for future research therefore centers on determining when such "simplistic" evaluation leads to inferential "errors" apart from the issues already considered in this paper.

(3) *Adjustment.* The contrast assumption concerning w_k in Eq. (1) is critical to the belief-adjustment model. As noted previously, a model that is almost identical to the SbS process when $R = 0$ has been proposed by Carlson and Dulany (1988). The purpose of their model is to explain how

people form judgments of the "convincingness" of evidence and then integrate this into an overall judgment recorded on a scale going from -1 to $+1$. Carlson and Dulany postulate that the convincingness of evidence is encoded positively or negatively vis-à-vis the hypothesis under consideration (i.e., $R = 0$) and that revision of belief is "proportional to remaining uncertainty." This latter assumption is mathematically equivalent to our contrast assumption but is not given the same rationale. Instead, it "reflects the intuition that completely convincing evidence should remove all uncertainty regardless of the absolute amount; and, by generalization, as the evidence is less convincing it should remove smaller proportions of uncertainty" (Carlson & Dulany, 1988, p. 468). In an experiment investigating the use of circumstantial evidence when reasoning in relation to a simulated crime scenario, Carlson and Dulany (1988) found strong support for their postulated process concerning the way people arrive at judgments of the convincingness of evidence. They also show that these judgments can be used to predict how people revise beliefs across time. As noted earlier, Ashton and Ashton (1988) formally tested and validated our contrast assumption (and Carlson and Dulany's "intuition") by measuring the size of adjustment when the same information was processed after different-sized anchors.

Limitations and Extensions

Our model is not a complete explanation of belief updating and order-effect phenomena and it suggests a number of directions for further theoretical and empirical work. One important and interesting topic centers on specifying how semantic content may interact with various task characteristics thereby influencing order effects in a manner unaccounted for in our predictions (see, e.g., Benassi, 1982). Another concerns how people deal with dependencies among different pieces of information. Imagine, for example, that symptom X is highly diagnostic of disease Y. However, if you knew that a person with X had already experienced the disease, the symptom might be irrelevant because of an acquired immunity. In our work, we have implicitly assumed that the outcomes of the coding process already include whatever conditioning the subject has done based on prior evidence. However, this should be studied in its own right. Similarly, Schum and Martin (1982) reported that when information processed by subjects (in a judicial context) was not in decomposed form (i.e., piece by piece), positive evidence was sometimes evaluated as negative (and vice versa). However, when subjects were given decomposed information that highlighted the structure of the data, fewer errors of this type were made. We note, parenthetically, that this issue is similar in structure to the "change of meaning" hypothesis that has been said to account for primacy effects in impression formation (Asch, 1946). Whereas our model shows that change of meaning is not necessary for primacy, it does seem likely to occur when there is ambiguity concerning the evidence to be evaluated (cf. Asch & Zukier, 1984).

Related to this issue is the problem of updating beliefs on the basis of new but redundant evidence. While we have not dealt with this issue, we note that Schum and Martin (1982) found that, "The most systematic result in our study concerns the holistic tendency to 'double-count' corroboratively redundant testimony" (p. 144). While such "double-counting" is consistent with our sequential anchoring-and-adjustment model because it is assumed that memory is limited to the location of one's current anchor and not how this was reached, much work will be necessary to understand how, and why, redundant information is integrated with existing beliefs.

Finally, whereas we have outlined a framework showing how task variables and processes interact in producing order effects in belief updating, much remains to be done in studying how further procedural and task variables affect this process. It would be illuminating, for example, to study the effects of variables such as time pressure, different forms of aggregated and disaggregated data presentation (see, e.g., Ashton & Ashton, 1988; Shanteau, 1970), temporal delays (see, e.g., Miller & Campbell, 1961; Insko, 1964; Luchins & Luchins, 1970; Kruglanski & Freund, 1983), the effects of interrupting updating tasks (see, e.g., Hoch, 1984; Luchins, 1957b), expertise (Messier & Tubbs, 1990), and how roles and incentives might affect differential sensitivity to negative and positive evidence.

For example, because the model distinguishes between sensitivity toward positive and negative evidence, it would be illuminating to investigate different combinations of the α and β parameters. These are illustrated graphically in Fig. 7 which shows how joint attitudes toward positive and negative evidence can be characterized by regions in the α, β space. The positions at the lower left ($\alpha = \beta = 0$) and upper right ($\alpha = \beta = 1$) corners represent, respectively, attitudes of extreme insensitivity and sensitivity to both negative and positive evidence. To illustrate, the parameters of someone with little or no prior information on a topic would be expected to lie in the upper right region. However, as information accumulates, the parameters of an unbiased person would be expected to move down toward the lower left region. The off-diagonal regions depict asymmetric attitudes toward positive and negative evidence. At one extreme, the lower right corner represents the position of the skeptic who is highly sensitive to negative evidence but ignores positive evidence ($\alpha = 1$, $\beta = 0$); at the other extreme, the top left corner represents the advocate who ignores negative evidence but is highly sensitive to positive evidence ($\alpha = 0$, $\beta = 1$).

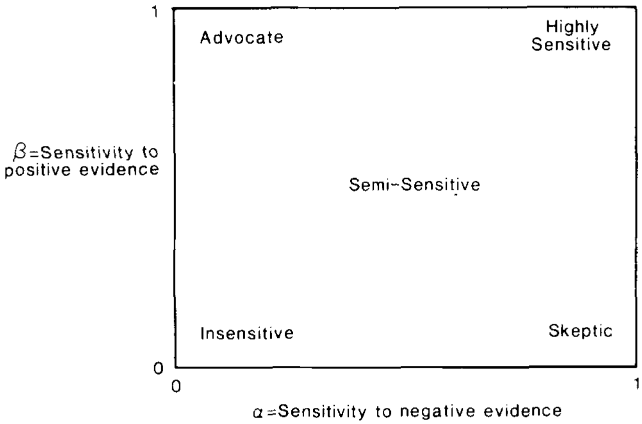

FIG. 7. Space of different possible attitudes toward positive and negative evidence.

Toward a Procedural Theory of Judgment

In recent years, the greatest challenge to those interested in decision making has been the extreme sensitivity of judgment and choice to seemingly minor changes in tasks (Einhorn & Hogarth, 1981; Payne, 1982). The importance and pervasiveness of various context effects (including framing, Tversky & Kahneman, 1981; response modes, Hogarth, 1982; and so on) may create a view of decision making as a fragmented and chaotic field. After all, if judgments and choices are sensitive to small changes in tasks, what hope is there of obtaining generalizable knowledge (cf. Cronbach, 1975)?

We believe that one answer lies in developing what Lopes (1982) has called a "procedural theory of judgment." That is, by focusing on the effects of task variables on information-processing strategies, complex behavior can be seen to arise from the interaction of simple psychological processes with an infinitely varied environment (see also Hogarth, 1986). Indeed, this general approach underlies our attempt to understand the updating of beliefs.

The updating of beliefs plays an important role in many areas of psychology. We have proposed and tested a model that illustrates how a simple psychological process, based on the sequential processing of information, can interact with task variables to produce a wide range of judgmental effects. We view our approach as providing a bridge between the idea that people are limited information processors, on the one hand, and the complexity of behavior and its sensitivity to environmental changes, on the other. Thus, our approach can be seen as an attempt to bring "chaos" (at the level of responses), out of "order" (in the underlying processes).

APPENDIX A

Classification of Order-Effect Studies by Task Variables (Details of Table 1)

Study and content	Complexity: Simple	Complexity: Complex	Length of series: Short	Length of series: Long	Response mode: EoS	Response mode: SbS	Order effects	Comments
1. Allen & Feldman (1974) Attributions of performance	x			x	x		Primacy	
2. Anderson (1959) Jury trial materials		x		x		x	Recency	Primacy at end of series
3. Anderson (1964) Number averaging	x		x			x	Recency	
4. Anderson (1965) Trait adjectives	x		x		x		Primacy	
5. Anderson (1967) Estimate average of 6 weights		x	x		x		Recency	
6. Anderson (1968) Trait adjectives	x		x			x	Recency	Adjectives of same polarity
	x		x		x		Recency	Subjects read each adjective aloud
								No significant effect in one condition
7. Anderson (1973a) Statesmanship of U.S. Presidents		x	x		x		Recency	
8. Anderson (1973b) Trait adjectives	x		x		x		Primacy	
9. Anderson & Barrios (1961) Trait adjectives	x		x		x		Primacy	
10. Anderson & Farkas (1973) Statesmanship of U.S. Presidents		x	x			x	Recency	
11. Anderson & Hubert (1963) Trait adjectives	x		x		x		Primacy	A recall condition reduced

12. Anderson & Jacobson (1968)								
Estimate average of 3 weights		x	x		x		Recency	
13. Anderson & Norman (1964)								
Trait adjectives	x		x		x		Primacy	
Foods	x		x		x		Primacy	
Headlines		x	x		x		Marginal primacy	
Life events		x	x		x		No effects	
14. Asch (1946)								
Trait adjectives	x		x		x		Primacy	
15. Benassi (1982)								
Attributions of performance	x			x	x		Primacy	Effects attenuated by attractiveness of those judged
16. Bossart & DiVesta (1966)								
Trait adjectives	x		x		x		Primacy	Adjectives were embedded in spoken messages
17. Chalmers (1971)								
Trait adjectives	x		x		x		Primacy	
18. Crano (1977)								
Contradictory messages		x	x		x		Recency	Subjects read two 600-word messages
19. Curley et al. (1988)								
Judgments of contingencies	x			x	x		Primacy	
20. Dale (1968)								
Probability judgments	x			x		x	Primacy	
21. Dreben, Fiske, & Hastie (1979)								
Impression formation	x		x		x		Primacy	
	x		x			x	Recency	
22. Feldman & Bernstein (1977)								
Attributions of own performance	x			x	x		Primacy	

APPENDIX A—*Continued*

Study and content	Complexity		Length of series		Response mode		Order effects	Comments
	Simple	Complex	Short	Long	EoS	SbS		
23. Feldstein & Bernstein (1978) Attributions of own performance	x			x	x		Primacy	
24. Furnham (1986) Mock trials		x	x			x	Recency	Judgments of guilt/innocence
25. Hendrick & Constantini (1970a) Trait adjectives	x		x		x		Primacy	
Trait adjectives	x		x		x		Recency	Subjects required to pronounce each adjective
26. Hendrick & Constantini (1970b) Number averaging	x		x			x	Recency	
27. Hendrick et al. (1973) Trait adjectives	x		x		x		Primacy	
28. Jones et al. (1968) Attributions of performance	x			x	x		Primacy	Concerning others (5 experiments)
	x			x	x		Recency	Concerning self (1 experiment)
29. Lana (1961) Attitudes toward vivisection		x		x	x		Primacy	Subjects unfamiliar with topic
		x		x	x		Recency	Subjects familiar with topic
30. Langer & Roth (1975) Outcomes of coin tosses	x			x	x		Primacy	
31. Levin (1976) Number averaging	x		x			x	Recency	
	x		x		x		No effect	

32. Levin & Schmidt (1969) Trait adjectives	x	x			x	Recency	
33. Levin & Schmidt (1970) Trait adjectives	x	x			x	Recency	
34. Levin et al. (1977) Ratings	x	x		x		Recency	Based on grades
35. Lichtenstein & Srull (1987) Behavior statements	x		x	x		Primacy	Under impression and memory sets
	x		x	x		Recency	Under a comprehension set
36. Luchins (1957a) Descriptions of people	x	x		x		Primacy	Traits embedded in text
37. Luchins (1957b) Descriptions of people	x	x		x		Primacy	
	x	x		x		Recency	Interpolation between two blocks of traits embedded in text
38. Luchins (1958) Descriptions of people	x	x			x	Recency	
39. Luchins & Luchins (1962a) Attributions of attitudes of another	x	x		x		Primacy	
40. Luchins & Luchins (1962b) Attributions of attitudes of another	x	x			x	Recency	
41. Luchins & Luchins (1970) Descriptions of people	x	x			x	Recency	

APPENDIX A—*Continued*

Study and content	Complexity		Length of series		Response mode		Order effects	Comments
	Simple	Complex	Short	Long	EoS	SbS		
42. Luchins & Luchins (1984) Descriptions of people	x		x			x	Recency	
43. Luchins & Luchins (1986) Judgments of moral/immoral behavior	x		x		x		Primacy	
44. McAndrew (1981) Attributions of performance	x			x	x		Primacy	
45. Newtson & Rindner (1979) Attributions of performance	x			x	x		Primacy	Effects attenuated when stimuli shown at high speed
46. Parducci et al. (1968) Estimate average loudness of sets of noise		x	x		x		Recency	
47. Pitz & Reinhold (1968) Probability judgments	x		x			x	Recency	
48. Pratz (1987) Attributions of own success	x			x	x		Primacy	
49. Roby (1967) Probability judgments	x			x		x	Primacy	
50. Rosenbaum & Levin (1968) Contradictory sentences		x	x		x		Recency	Subjects told to expect two lists of adjectives
51. Shanteau (1970) Probability judgments	x		x			x	Recency	Stimuli were red and white lights

52. Shanteau (1972)								
Probability judgments	x		x			x	Recency	Stimuli were pink and white beads
53. Stewart (1965)								
Trait adjectives	x		x		x		Primacy	
Trait adjectives	x		x			x	Recency	
54. Strange et al. (1978)								
Trait adjectives	x		x		x		Primacy	
	x		x		x		Recency	Interpolation (see text)
	x		x		x		No effect	Interpolation (see text)
55. Tetlock (1983)								
Simulated legal trial evidence		x		x	x		Primacy	
		x		x	x		No effect	Instructions to justify responses
56. Tesser (1968)								
Trait adjectives	x		x		x		Primacy	
57. Tiwari (1978)								
Trait adjectives	x		x		x		No effect	
58. Walker et al. (1972)								
Mock trial		x		x		x	Recency	
59. Weiss & Anderson (1969)								
Estimate average lengths of lines		x	x		x		Recency	
		x	x			x	Recency	
60. Yates & Curley (1986)								
Judgments of contingencies	x			x	x		Primacy	

Note. In establishing Table 1, each line of this appendix was treated as a single data point. Although studies often included more than one experiment, we have treated them as single data points except when the task conditions used in our classification scheme were explicitly varied. We have not included in our analysis the early studies conducted by Lund (1925), Cromwell (1950), and Hovland & Mandell (1957).

APPENDIX B

Order Effects for the SbS Process when $R = S_{k-1}$

As stated in the text, order effects for the SbS process when $R = S_{k-1}$ always imply recency effects. We demonstrate the conditions under which this statement holds following the proof provided in Anderson and Hovland (1957).

Consider two evidence items, a and b, with subjective strengths $s(x_a)$ and $s(x_b)$ where, without loss of generality, $s(x_a) < s(x_b)$. Define opinion reached after processing the evidence in the order a followed by b by S_{ab}, and opinion reached after processing the same information in the reverse order by S_{ba}. Let the order effect be defined by $D = S_{ab} - S_{ba}$ and note that recency is implied by $D > 0$. By successively applying Eq. (1) in the text, we can write

$$S_{ab} = S_0 + w_a[s(x_a) - S_0] + w_{ab}[s(x_b) - S_0 - w_a\{s(x_a) - S_0\}] \tag{B.1}$$

and

$$S_{ba} = S_0 + w_b[s(x_b) - S_0] + w_{ba}[s(x_a) - S_0 - w_b\{s(x_b) - S_0\}], \tag{B.2}$$

where w_a and w_{ba} refer to the adjustment weights when evidence item a is processed first and second, respectively, and w_b and w_{ab} are the analogous adjustment weights in respect of evidence item b ($0 \leq w_a, w_{ba}, w_b, w_{ab} \leq 1$). By subtracting Eq. (B.2) from (B.1), we obtain

$$D = [s(x_a) - S_0][w_a - w_a w_{ab} - w_{ba}] + [s(x_b) - S_0][w_{ab} - w_b + w_b w_{ba}]. \tag{B.3}$$

Anderson and Hovland (1957) assume that $w_a = w_{ba}$ and $w_b = w_{ab}$ in which case Eq. (B.3) can be reexpressed as

$$D = [s(x_b) - S_0]w_a w_b - [s(x_a) - S_0]w_a w_b \tag{B.4}$$

or

$$D = w_a w_b[s(x_b) - s(x_a)]. \tag{B.5}$$

Because $s(x_b) > s(x_a)$, $D > 0$ and recency always obtains.

We note that in our model it is not necessarily the case that $w_a = w_{ba}$ and that $w_b = w_{ab}$. However, this assumption considerably simplifies the algebraic proof of recency effects. In addition, the form of Eq. (B.5) shows that the size of recency effects is proportional to the difference between $s(x_b)$ and $s(x_a)$.

APPENDIX C

Order Effects for the SbS Process when $R = 0$

Consistent evidence. For consistent evidence, no order effects are predicted by the SbS process when $R = 0$. To see why in the case of negative evidence $[s(x_k) < 0]$, note that Eq. (7a) in the text can be rewritten in the form

$$S_k = S_{k-1}\,[1 + \alpha\, s(x_k)]. \qquad \text{(C.1)}$$

Thus when $k = 2$, Eq. (C.1) can be written $S_2 = S_1\,[1 + \alpha\, s(x_2)]$ and further expanded as

$$S_2 = S_0\,[1 + \alpha\, s(x_1)][1 + \alpha\, s(x_2)]. \qquad \text{(C.2)}$$

Since multiplication is commutative, S_2 is not affected by the order in which $s(x_1)$ and $s(x_2)$ are processed, a result that can be generalized for $k > 2$. Thus, the model predicts no order effects.

In analogous fashion, the same prediction can be made for the case of consistent positive evidence. Specifically, Eq. (7b) in the text can be rewritten as

$$S_k = S_{k-1}\,[1 - \beta\, s(x_k)] + \beta\, s(x_k), \qquad \text{(C.3)}$$

which, when $k = 2$, can be expanded to the form

$$\begin{aligned} S_2 = S_0 + \beta(1 - S_0)\,[s(x_1) \\ + s(x_2) - \beta\, s(x_1)s(x_2)]. \end{aligned} \qquad \text{(C.4)}$$

Since both addition and multiplication are commutative, it can be seen that the order in which $s(x_1)$ and $s(x_2)$ are processed has no effect on S_2, a result that can also be generalized for $k > 2$.

Mixed evidence. When processing a sequence of mixed evidence the SbS process (with $R = 0$) does lead to order effects, and specifically recency.

Consider the effects of processing a negative piece of evidence, denoted $s(x_-)$, followed by a positive piece, denoted $s(x_+)$, and then the positive followed by the negative. Define an order effect as

$$D = S(-, +) - S(+, -). \qquad \text{(C.5)}$$

This can be written

$$\begin{aligned} D &= [S_0 + w_1\, s(x_-) + r_2\, s(x_+)] - [S_0 + r_1 s(x_+) + w_2\, s(x_-)] \\ &= s(x_+)(r_2 - r_1) + s(x_-)(w_1 - w_2). \end{aligned} \qquad \text{(C.6)}$$

By definition, $r_1 = \beta\,(1 - S_0)$ and $r_2 = \beta[1 - \{S_0 + w_1 s(x_-)\}]$ such that $(r_2 - r_1) = -\beta\, w_1 s(x_-)$ or $-S_0\,\alpha\,\beta\, s(x_-)$. Similarly, $w_1 = \alpha\, S_0$ and $w_2 = \alpha[S_0 + r_1 s(x_+)]$ such that $(w_1 - w_2) = -\alpha\, r_1 s(x_+)$ or $-(1 - S_0)\,\alpha\,\beta\, s(x_+)$. Substituting these values into Eq. (C.6) we obtain

$$\begin{aligned} D &= -s(x_+)S_0\,\alpha\,\beta\, s(x_-) - s(x_-)(1 - S_0)\alpha\,\beta\, s(x_+) \\ &= -\alpha\,\beta\, s(x_-)\, s(x_+) > 0, \end{aligned} \tag{C.7}$$

because $s(x_-) < 0$.

Thus for two pieces of mixed evidence, the recency effect is the product of both sensitivity to evidence (α and β) and the subjective strengths of the two pieces of evidence. Note also that the model predicts that the magnitude of the recency effect is unrelated to the size of the initial anchor (an implication that was explicitly tested and verified by Ashton and Ashton, 1988).

REFERENCES

Allen, V. L., & Feldman, R. S. (1974). Tutor attributions and attitudes as a function of tutee performance. *Journal of Applied Social Psychology,* **4,** 311–320.

Anderson, N. H. (1959). Test of a model for opinion change. *Journal of Abnormal and Social Psychology,* **59,** 371–381.

Anderson, N. H. (1964). Test of a model for number-averaging behavior. *Psychonomic Science,* **1,** 191–192.

Anderson, N. H. (1965). Primacy effects in personality impression formation using a generalized order effect paradigm. *Journal of Personality and Social Psychology,* **2,** 1–9.

Anderson, N. H. (1967). Application of a weighted average model to a psychophysical averaging task. *Psychonomic Science,* **8,** 227–228.

Anderson, N. H. (1968). Application of a linear-serial model to a personality-impression task using serial presentation. *Journal of Personality and Social Psychology,* **10,** 354–362.

Anderson, N. H. (1973a). Information integration theory applied to attitudes about U.S. presidents. *Journal of Educational Psychology,* **64,** 1–8.

Anderson, N. H. (1973b). Serial position curves in impression formation. *Journal of Experimental Psychology,* **97,** 8–12.

Anderson, N. H. (1981). *Foundations of information integration theory.* New York: Academic Press.

Anderson, N. H., & Barrios, A. A. (1961). Primacy effects in personality impression formation. *Journal of Abnormal and Social Psychology,* **63,** 346–350.

Anderson, N. H., & Farkas, A. J. (1973). New light on order effects in attitude change. *Journal of Personality and Social Psychology,* **28,** 88–93.

Anderson, N. H., & Hovland, C. I. (1957). In C. I. Hovland (Ed.), *The order of presentation in persuasion* (pp. 158–169). New Haven, CT: Yale Univ. Press.

Anderson, N. H., & Hubert, S. (1963). Effects of concomitant verbal recall on order effects in personality impression formation. *Journal of Verbal Learning and Verbal Behavior,* **2,** 379–391.

Anderson, N. H., & Jacobson, A. (1968). Further data on a weighted average model for judgment in a lifted weight task. *Perception & Psychophysics,* **4,** 81–84.

Anderson, N. H., & Norman, A. (1964). Order effects in impression formation in four classes of stimuli. *Journal of Abnormal and Social Psychology,* **69,** 467–471.

Asare, S. (1989). *The auditor's going concern decision: Interaction of task variables and the sequential processing of evidence.* Unpublished manuscript, Fisher School of Accounting, University of Florida, Gainesville.

Asch, S. E. (1946). Forming impressions of personality. *Journal of Abnormal and Social Psychology* , **41**, 258–290.

Asch, S. E., & Zukier, H. (1984). Thinking about persons. *Journal of Personality and Social Psychology,* **46**, 1230–1240.

Ashton, A. H., & Ashton, R. H. (1988). Sequential belief revision in auditing. *The Accounting Review,* **63**, 623–641.

Benassi, M. (1982). Effects of order presentation, primacy, and physical attractiveness on attributions of ability. *Journal of Personality and Social Psychology,* **43**, 48–58.

Bossart, P., & DiVesta, F. J. (1966). Effects of context, frequency, and order of presentation of evaluative assertions on impression formation. *Journal of Personality and Social Psychology,* **4**, 538–544.

Brunswik, E. (1952). *The conceptual framework of psychology.* Chicago: Univ. of Chicago Press.

Butt, J. L., & Campbell, T. L. (1989). The effects of information order and hypothesis testing strategies on auditors' judgments. *Accounting, Organizations and Society,* **14**, 471–479.

Camerer, C. F. (1987). Do biases in probability judgments matter in markets? Experimental evidence. *American Economic Review,* **77**, 981–987.

Carlson, R. A., & Dulany, D. E. (1988). Diagnostic reasoning with circumstantial evidence. *Cognitive Psychology,* **20**(4), 463–492.

Chalmers, D. K. (1971). Repetition and order effects in attitude formation. *Journal of Personality and Social Psychology,* **17**, 219–228.

Cooper, J., & Croyle, R. T. (1984). Attitudes and attitude change. *Annual Review of Psychology,* **35**, 395–426.

Crano, W. D. (1977). Primacy versus recency in retention of information and opinion change. *Journal of Social Psychology,* **101**, 87–96.

Cromwell, H. (1950). The relative effect on audience attitude of the first versus the second argumentative speech of a series. *Speech Monographs,* **17**, 105–122.

Cronbach, L. J. (1975). Beyond the two disciplines of scientific psychology. *American Psychologist,* **30**, 116–127.

Curley, S. P., Young, M. J., Kingry, M. J., & Yates, J. F. (1988). Primacy effects in clinical judgments of contingency. *Medical Decision Making,* **8**, 216–222.

Dale, H. C. A. (1968). Weighing evidence: An attempt to assess the efficiency of the human operator. *Ergonomics,* **11**, 215–230.

Davis, J. H. (1984). Order in the courtroom. In D. J. Miller, D. G. Blackman, & A. J. Chapman (Eds.), *Perspectives in psychology and law.* New York: Wiley.

De Wit, H., Hogarth, R. M., Koehler, J. J., & Luchins, D. J. (1989). Effects of diazepam on a belief-updating task. *Psychological Reports,* **64**, 219–226.

Dillard, J. F., Kauffman, N. L., & Spires, E. E. (in press). Evidence order and belief revision in management accounting decisions. *Accounting, Organizations and Society.*

Dreben, E. K., Fiske, S. T., & Hastie, R. (1979). The independence of evaluative and item information: Impression and recall order effects in behavior-based impression formation. *Journal of Personality and Social Psychology,* **37**, 1758–1768.

Edwards, W. (1968). Conservatism in human information processing. In B. Kleinmuntz (Ed.), *Formal representation of human judgment.* New York: Wiley.

Einhorn, H. J., & Hogarth, R. M. (1981). Behavioral decision theory: Processes of judgment and choice. *Annual Review of Psychology,* **32**, 53–88.

Einhorn, H. J., & Hogarth, R. M. (1986). Judging probable cause. *Psychological Bulletin,* **99,** 3–19.

Feldman, R. S., & Bernstein, A. G. (1977). Degree and sequence of success as determinants of self-attribution of ability. *The Journal of Social Psychology,* **102,** 223–231.

Feldman, R. S., & Bernstein, A. G. (1978). Primacy effects in self-attribution of ability. *Journal of Personality,* **46,** 732–742.

Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention and behavior: An introduction to theory and research.* Reading, MA: Addison–Wesley.

Fischhoff, B., & Beyth-Marom, R. (1983). Hypothesis evaluation from a Bayesian perspective. *Psychological Review,* **90,** 239–260.

Furnham, A. (1986). The robustness of the recency effect: Studies using legal evidence. *The Journal of General Psychology,* **113,** 351–357.

Goldstein, W. M., & Einhorn, H. J. (1987). Expression theory and the preference reversal phenomena. *Psychological Review,* **94,** 236–254.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics.* New York: Wiley.

Hastie, R. (1983). Social inference. *Annual Review of Psychology,* **34,** 511–542.

Hatvany, N., & Strack, F. (1980). The impact of a discredited key witness. *Journal of Applied Social Psychology,* **10,** 490–509.

Hendrick, C., & Constantini, A. F. (1970a). Effects of varying trait inconsistency and response requirements on the primacy effect in impression formation. *Journal of Personality and Social Psychology,* **15,** 158–164.

Hendrick, C., & Constantini, A. F. (1970b). Number averaging behavior: A primacy effect. *Psychonomic Science,* **19,** 121–122.

Hendrick, C., Constantini, A. F., McGarry, J., & McBride, K. (1973). Attention decrement, temporal variation, and the primacy effect in impression formation. *Memory & Cognition,* **1,** 193–195.

Hoch, S. J. (1984). Availability and interference in predictive judgment. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* **10,** 649–662.

Hogarth, R. M. (1975). Cognitive processes and the assessment of subjective probability distributions. *Journal of the American Statistical Association,* **70,** 271–289.

Hogarth, R. M. (Ed.). (1982). *New directions for methodology of behavioral sciences: Question framing and response consistency.* San Francisco: Jossey-Bass.

Hogarth, R. M. (1986). Generalization in decision research: The role of formal models. *IEEE Transactions on Systems, Man, and Cybernetics, SMC*-16(3), 439–449.

Hovland, C. I., Janis, I. L., & Kelley, H. H. (1953). *Communication and persuasion: Psychological studies of opinion change.* New Haven, CT: Yale Univ. Press.

Hovland, C. I., & Mandell, W. (1957). Is there a "law of primacy in persuasion?" In C. I. Hovland (Ed.), *The order of presentation in persuasion* (pp. 13–22). New Haven, CT: Yale Univ. Press.

Insko, C. A. (1964). Primacy versus recency in persuasion as a function of the timing of arguments and measures. *Journal of Abnormal and Social Psychology,* **69,** 381–391.

Jones, E. E. (1979). The rocky road from acts to dispositions. *American Psychologist,* **34,** 104–117.

Jones, E. E., & Goethals, G. R. (1972). Order effects in impression formation: Attribution context and the nature of the entity. In E. E. Jones et al. (Eds.), *Attribution: Perceiving the causes of behavior.* Morristown, NJ: General Learning Press.

Jones, E. E., Rock, L., Shaver, K. G., Goethals, G. R., & Ward, L. M. (1968). Pattern of performance and ability attribution. *Journal of Personality and Social Psychology,* **10,** 317–340.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, **47**, 263–291.

Koch, B. S., Pei, B. K. W., & Reed, S. A. (1989, July). *Auditor belief revisions in a performance auditing setting: An application of the contrast-inertia model.* Unpublished manuscript, The University of North Texas, Denton, TX.

Kruglanski, A. W., & Freund, T. (1983). The freezing and unfreezing of lay-inferences: Effects on impressional primacy, ethnic stereotyping, and numerical anchoring. *Journal of Experimental Social Psychology*, **19**, 448–468.

Lana, R. E. (1961). Familiarity and the order of presentation of persuasive communications. *Journal of Abnormal and Social Psychology*, **62**, 573–577.

Langer, E. J., & Roth, J. (1975). Heads I win, tails it's chance: The illusion of control as a function of the sequence of outcomes in a purely chance task. *Journal of Personality and Social Psychology*, **32**, 951–955.

Levin, I. P. (1976). Processing of deviant information in inference and descriptive tasks with simultaneous and serial presentation. *Organizational Behavior and Human Performance*, **15**, 195–211.

Levin, I. P., Ims, J. R., Simpson, J. C., & Kim, K. J. (1977). The processing of deviant information in prediction and evaluation. *Memory & Cognition*, **5**, 679–684.

Levin, I. P., & Schmidt, C. F. (1969). Sequential effects in impression formation with binary intermittent responding. *Journal of Experimental Psychology*, **79**, 283–287.

Levin, I. P., & Schmidt, C. F. (1970). Differential influence of information in an impression-formation task with binary intermittent responding. *Journal of Experimental Psychology*, **84**, 374–376.

Lichtenstein, M., & Srull, T. K. (1987). Processing objectives as a determinant of the relationship between recall and judgment. *Journal of Experimental Social Psychology*, **23**, 93–118.

Lopes, L. L. (1982). *Toward a procedural theory of judgment.* University of Wisconsin, Department of Psychology, Wisconsin Information Processing Program.

Lopes, L. L. (1985). Averaging rules and adjustment processes in Bayesian influence. *Bulletin of the Psychonomic Society*, **23**(6), 509–512.

Lopes, L. L. (1987a). Procedural debiasing. *Acta Psychologica*, **64**(2), 167–185.

Lopes, L. L. (1987b). Beyond hope and fear: The psychology of risk. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology* (Vol. 20, pp. 255–295). New York: Academic Press.

Luchins, A. S. (1957a). Primacy-recency in impression formation. In C. I. Hovland (Ed.), *The order of presentation in persuasion* (pp. 33–61). New Haven, CT: Yale Univ. Press.

Luchins, A. S. (1957b). Experimental attempts to minimize the impact of first impressions. In C. I. Hovland (Ed.), *The order of presentation in persuasion* (pp. 62–75). New Haven, CT: Yale Univ. Press.

Luchins, A. S. (1958). Definitiveness of impression and primacy-recency in communications. *The Journal of Social Psychology*, **48**, 275–290.

Luchins, A. S., & Luchins, E. H. (1962a). Personality impressions from communications reflecting attitudes toward segregation. *The Journal of Social Psychology*, **58**, 315–330.

Luchins, A. S., & Luchins, E. H. (1962b). Primacy-recency in communications reflecting attitudes toward segregation. *The Journal of Social Psychology*, **58**, 357–369.

Luchins, A. S., & Luchins, E. H. (1970). The effects of order of presentation of information and explanatory models. *The Journal of Social Psychology*, **80**, 63–70.

Luchins, A. S., & Luchins, E. H. (1984). Conceptions of personality and order effects in forming impressions. *The Journal of General Psychology*, **110**, 165–196.

Luchins, A. S., & Luchins, E. H. (1986). Primacy and recency effects with descriptions of moral and immoral behavior. *The Journal of General Psychology*, **113**, 159–177.

Lund, F. H. (1925). The psychology of belief: A study of its emotional and volitional determinants. *Journal of Abnormal Psychology,* **20,** 174–196.

McAndrew, F. T. (1981). Pattern of performance and attributions of ability and gender. *Personality and Social Psychology Bulletin,* **7,** 583–587.

Messier, W. F., Jr. (1989, December). *The sequencing of audit evidence: Its impact on the extent of audit testing and report formulation.* Unpublished manuscript, Fisher School of Accounting, University of Florida, Gainesville.

Messier, W. F., Jr., & Tubbs, R. M. (1990, February). *Mitigating recency effects in belief revision: The impact of experience and review.* Unpublished manuscript, Fisher School of Accounting, University of Florida, Gainesville.

Miller, N., & Campbell, D. T. (1961). Recency and primacy in persuasion as a function of the timing of speeches and measurement. *Journal of Abnormal and Social Psychology,* **59,** 1–9.

Newell, A., & Simon, H. A. (1972). *Human problem solving.* Englewood Cliffs, NJ: Prentice–Hall.

Newtson, D., & Rindner, R. J. (1979). Variation in behavior perception and ability attribution. *Journal of Personality and Social Psychology,* **37,** 1847–1858.

Nisbett, R., & Ross, L. (1980). *Human inference: Strategies and shortcomings of human judgment.* Englewood Cliffs, NJ: Prentice–Hall.

Parducci, A., Thaler, H., & Anderson, N. H. (1968). Stimulus averaging and the context for judgment. *Perception and Psychophysics,* **3,** 145–150.

Payne, J. W. (1982). Contingent decision behavior. *Psychological Bulletin,* **92,** 382–402.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1990). The adaptive decision-maker: Effort and accuracy in choice. In R. M. Hogarth (Ed.), *Insights in decision making* (pp. 129–153). Chicago, IL: Univ. of Chicago Press.

Pennington, N., & Hastie, R. (1986). Evidence evaluation in complex decision making. *Journal of Personality and Social Psychology,* **51,** 242–258.

Peterson, C. R., & Beach, L. R. (1967). Man as an intuitive statistician. *Psychological Bulletin,* **68,** 29–46.

Peterson, C. R., & DuCharme, W. M. (1967). A primacy effect in subjective probability revision. *Journal of Experimental Psychology,* **73,** 61–65.

Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 19, pp. 123–205). New York: Academic Press.

Pitz, G. F., & Reinhold, H. (1968). Payoff effects in sequential decision-making. *Journal of Experimental Psychology,* **77,** 249–257.

Pratz, O. (1987). Computer presentation of pattern-of-success: A primacy effect verified. *Perceptual and Motor Skills,* **65,** 379–387.

Raiffa, H. A., & Schlaifer, R. (1961). *Applied statistical decision theory.* Cambridge, MA: MIT Press.

Robinson, L. B., & Hastie, R. (1985). Revision of beliefs when a hypothesis is eliminated from consideration. *Journal of Experimental Psychology: Human Perception and Performance,* **11,** 443–456.

Roby, T. B. (1967). Belief states and sequential evidence. *Journal of Experimental Psychology,* **75,** 236–245.

Rosenbaum, M. E., & Levin, I. P. (1968). Impression formation as a function of source credibility and order of presentation of contradictory information. *Journal of Personality and Social Psychology,* **10,** 167–174.

Ross, L., & Lepper, M. R. (1980). The perseverance of beliefs: Empirical and normative considerations. In R. A. Shweder (Ed.), *New directions for methodology of behavioral sciences: Fallible judgment in behavioral research.* San Francisco: Jossey-Bass.

Schum, D. A. (1980). Current developments in cascaded inference. In T. S. Wallsten (Ed.), *Cognitive processes in choice and decision behavior.* Hillsdale, NJ: Erlbaum.

Schum, D. A., & Martin, A. W. (1980, October). *Probabilistic opinion revision on the basis of evidence at trial: A Baconian or a Pascalian process?* (Report No. 80-02). Houston, TX: Rice University.

Schum, D. A., & Martin, A. W. (1982). Formal and empirical research on cascaded inference in jurisprudence. *Law and Society Review,* **17,** 105–151.

Shafer, G. (1976). *A mathematical theory of evidence.* Princeton, NJ: Princeton Univ. Press.

Shanteau, J. C. (1970). An additive model for sequential decision making. *Journal of Experimental Psychology,* **85,** 181–191.

Shanteau, J. C. (1972). Descriptive versus normative models of sequential inference judgment. *Journal of Experimental Psychology,* **93,** 63–68.

Simon, H. A., & Newell, A. (1971). Human problem solving: The state of the theory in 1970. *American Psychologist,* **26,** 145–159.

Slovic, P., & Lichtenstein, S. (1971). Comparison of Bayesian and regression approaches to the study of information processing in judgment. *Organizational Behavior and Human Performance,* **6,** 649–744.

Stewart, R. H. (1965). Effect of continuous responding on the order effect in personality impression formation. *Journal of Personality and Social Psychology,* **1,** 161–165.

Strange, K. R., Schwei, M., & Geiselman, R. E. (1978). Effects of the structure of descriptions on group impression formation. *Bulletin of the Psychonomic Society,* **12,** 224–226.

Tesser, A. (1968). Differential weighting and directed meaning as explanations of primacy in impression formation. *Psychonomic Science,* **11,** 299–300.

Tetlock, P. E. (1983). Accountability and the perseverance of first impressions. *Social Psychology Quarterly,* **46,** 285–292.

Tiwari, B. D. (1978). Primacy-recency effects in impression-formation as a function of dependence-independence proneness. *Psychologia,* **21,** 143–149.

Triandis, H. C. (1971). *Attitude and attitude change.* New York: Wiley.

Tubbs, R. M., Messier, W. F., Jr., & Knechel, W. R. (1990). Recency effects in the auditor's belief revision process. *The Accounting Review,* **65,** 452–460.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science,* **211,** 453–458.

Tversky, A., Sattath, S., & Slovic, P. (1988). Contingent weighting in judgment and choice. *Psychological Review,* **95,** 371–384.

Walker, L., Thibaut, J., & Andreoli, V. (1972). Order of presentation at trial. *Yale Law Journal,* **82,** 216–226.

Weiss, D. J., & Anderson, N. H. (1969). Subjective averaging of length with serial presentation. *Journal of Experimental Psychology,* **82,** 52–63.

Winkler, R. L. (1972). *Introduction to Bayesian inference and decision.* New York: Holt, Rinehart & Winston.

Yates, J. F., & Curley, S. P. (1986). Contingency judgment: Primacy effects and attention decrement. *Acta Psychologica,* **62,** 293–302.

(Accepted January 21, 1991)