The Impact of Sequential Data on Consumer Confidence in Relative Judgments

DIPAYAN BISWAS
GUANGZHI ZHAO
DONALD R. LEHMANN

We examine how consumers update their confidences in ordinal (relative) judgments while evaluating sequential product-ranking and source-accuracy data in percentage versus frequency formats. The results show that when sequential data are relatively easier to mathematically combine (e.g., percentage data), consumers revise their judgments in a way that is consistent with an averaging model but inconsistent with the normative Bayesian model. However, when the sequential data are difficult to mathematically combine (e.g., frequency data), consumers update their confidence judgments in a way that is more consistent with the normative Bayesian model than with an averaging model. Interestingly, greater processing motivation for sequential frequency data leads to updated confidence judgments that are lower than normative Bayesian predictions but consistent with the averaging model. Overall, the results of the experiments reveal counterintuitive findings; updated confidence judgments are higher and more accurate when sequential data are more difficult to process and also when consumers have lower processing motivation.

Product-ranking and source-accuracy information is ubiquitous in the form of stock rankings by financial institutions, movie rankings by film critics, product rankings by consumer agents, product safety rankings by government agencies, rankings of financial analysts based on past accuracy, pollster accuracy rankings, and product rankings posted online by Web sites and individual consumers (Bialik 2006; Gershoff, Broniarczyk, and West 2001; Jadad and Gagliardi 1998; Kichen and Ray 2009). For example, financial analysts and business publications (e.g., *Business Week*, *Barron's*, *Forbes*) often provide rankings of mutual funds in terms of likely future performance and rankings of analysts based on their past accuracy levels (Kichen and Ray 2009; Simons 1998; Vinod 2004). Similarly, *Consumer Reports* and other publications (e.g., *US News and World Report*) provide rankings of products like cars and computers on various attributes.

Consumers themselves frequently make ordinal or relative judgments whereby they rank products in terms of their relative performances—for example, "Car B is safer than Car A" (Fox and Levav 2000). A crucial aspect of these judgments is the consumer's confidence in the relative judgment (e.g., Bearden, Hardesty, and Rose 2001)—that is, their self-assessed likelihood that their rank ordering is actually correct. Confidence in relative judgments is critical in determining product choices and the prices consumers are willing to pay. For instance, a consumer's degree of confidence in her relative judgment that Honda is better than Nissan on the attributes (e.g., safety, reliability) important to her will influence her willingness to pay a higher price for a Honda, as will her decision on whether to buy or to defer her choice. Similarly, an investor's decision regarding

Dipayan Biswas is associate professor of marketing, Bentley University, 175 Forest Street, Waltham, MA 02452 (dbiswas@bentley.edu). Guangzhi Zhao is assistant professor of marketing, University of Kansas, 1300 Sunnyside Avenue, Lawrence, KS 66045 (zhaog@ku.edu). Donald R. Lehmann is George E. Warren Professor of Business, Columbia University, 3022 Broadway, New York, NY 10027 (drl2@columbia.edu). Work on this research was partially supported by the Bentley Faculty Development Grant and by the University of Kansas General Research Fund. The authors thank Paul Berger, Ashok Lalwani, Bruce Weinberg, and the *JCR* editor, associate editor, and reviewers for helpful comments and suggestions.

Ann McGill served as editor and Brian Ratchford served as associate editor for this article.

Electronically published August 24, 2010

 0093-5301/2011/3705-0004$10.00. DOI: 10.1086/656061

allocation of funds to different financial products is likely to be influenced by the investor's degree of confidence regarding the relative performances of the financial products.

As an illustrative example, suppose a college professor is considering two comparable retirement funds (say, TIAA-CREF and Fidelity) for her retirement contributions (of, say, $1,000 per month), and her allocation of money to these two funds is driven by the relative likely future return potential of each fund. She refers to two different sources (say, *Morningstar* and *Zacks*) to get their opinions on these two funds. Assume the investor (i.e., the professor) first learns that *Morningstar* ranks Fidelity better than TIAA-CREF in terms of likely returns and that *Morningstar* has been correct 80% of the time with similar past predictions. Later, the investor learns that a different source, *Zacks*, also ranks Fidelity as better and that *Zacks* has been correct 70% of the time in past predictions. Based on this, the investor is expected to believe that Fidelity will perform better than TIAA-CREF since both agents rank Fidelity as better. However, from past literature, it is unclear how confident the investor will be that this relative (ordinal) judgment is correct. If this investor simply calculates the average accuracy of the two sources (e.g., Dougherty and Shanteau 1999), she will be 75% confident that Fidelity will perform better than TIAA-CREF. Based on this, she might allocate $750 to Fidelity and $250 to TIAA-CREF. However, since the two sources are independent, the investor's updated confidence should be greater than 80%. In fact, if the investor follows a normative Bayesian model (e.g., McKenzie 1994), she would be 90.32% confident in her relative judgment that Fidelity is likely to perform better than TIAA-CREF; hence, she would allocate $903 to Fidelity and $97 to TIAA-CREF (the exact calculations for these will be shown in eqq. 1 and 2 below).

Prior research has generally focused on cardinal ratings, and comparatively little work has been done on relative/ordinal judgments, in which one of two options is better (Fox and Levav [2000] being a notable exception). Here we examine how consumers update (i.e., change) their confidences in relative judgments with sequential data. Although consumer confidence in relative judgments can be influenced by several factors (e.g., product knowledge, goals, time pressure), we focus on the effects of sequential reception of product-ranking data from different agents with given past accuracy rates. In addition, we examine the role of the data format (percentage vs. frequency) and processing motivation on the updating of confidence in relative judgments.

Confidence in judgments (e.g., attitudes) has long been recognized as an important determinant of behavior (Bennett and Harrell 1975; Howard and Sheth 1969). While there has been extensive research on how confidence judgments are formed, these studies focused on issues such as individuals' overconfidence and underconfidence, confidence in their own general knowledge, and confidence in judgment of single versus multiple events (e.g., Gigerenzer, Hoffrage, and Kleinbölting 1991; Juslin and Olsson 1997; Kruger and Dunning 1999). Also, prior studies have relied on memory-based features of the confidence "formation process in which pieces of information are recalled from memory and integrated to form [the confidence] judgment" (Jacoby et al. 2002, 31). On the other hand, hardly any research has examined how confidence in relative judgments is updated as information is acquired sequentially.

Drawing on past research (Garafalo and Lester 1985; Gollwitzer and Schaal 1998; Schwarz 2004), we propose that when sequential data are presented in an easy-to-compute format (e.g., in percentage format, as in "*Fortune* has been correct 70% out of 177 times in its mutual fund rankings"), consumer confidence in relative judgment is likely to be revised by averaging the sequential data. However, for sequential data presented in a difficult-to-compute format (e.g., in frequency format, as in "*Fortune* has been correct 124 out of 177 times"), consumer confidence updating will be in a similar direction as the normative Bayesian model.

As will be discussed in detail later, the findings of our research have implications for the literature on judgment updating and especially the degree to which updated judgments are consistent with the normative Bayesian model. Specifically, prior research makes mixed claims or assumptions regarding the extent to which consumers update their judgments in a way similar to the Bayesian model. Our studies show that the degree of judgment updating is influenced by the perceived difficulty and motivation in processing sequential data. Interestingly, our experiments also show that simplifying heuristics (rather than systematic mathematical processing) while evaluating sequential data can lead to improved (i.e., normatively correct) updated judgments. Along with the theoretical implications, a key contribution of the research is in the empirical findings that have practical implications. For instance, the surgeon general's 2004 report *The Health Consequences of Smoking: A Report of the Surgeon General* (www.surgeongeneral.gov/library/smokingconsequences) presents sequential data in frequency format (across different pages) regarding different types of cancers caused by smoking, such as "An estimated 171,900 new cases of lung cancer were expected to be diagnosed . . . and an estimated 157,200 deaths attributable to lung cancer were expected to occur," and "An estimated 22,400 new cases and 12,100 deaths from cancer of the stomach were expected to occur." The surgeon general's report could have presented the same data in percentage format instead, such as "91% deaths out of 171,900 new cases" and "54% deaths out of 22,400 new cases." Which of these formats (frequency vs. percentage) would lead to normatively more accurate updated confidence judgments? The results of our studies show that the frequency format is more likely to lead to updated confidence judgments that are closer to the normative (Bayesian) level.

We examine updating of confidence in relative judgments in four studies. Study 1 demonstrates that consumers update their confidences in relative judgments to a greater extent when the data are presented in frequency rather than percentage format. Study 2 replicates and extends the findings of study 1 by directly manipulating the degree of difficulty in computing sequential frequency data. Study 3 extends the findings of studies 1 and 2 by examining the moderating effects of consumer processing motivation. Finally, study 4 rules out an alternative explanation for the key findings of studies 1–3 by examining the effects of highlighting the independence of the sequential data. We report robust findings across a wide range of products, such as financial products (studies 1 and 3), cars (study 2), and laptops (study 4).

THEORETICAL BACKGROUND

Consumers are expected to revise their confidences in relative judgments as they acquire additional information. Although no research has examined how consumer confidence in relative judgment is revised with sequential data in different (percentage vs. frequency) formats, there has been a considerable amount of research on how individuals revise and update their beliefs and judgments with new information (e.g., Anderson 1981; Edwards 1968; Hogarth and Einhorn 1992; Rust et al. 1999; Shanteau 1975; Slovic and Lichtenstein 1971). Two prominent models of individual information integration have been the Bayesian model and the averaging model.

Bayesian-Style Updating

One way of adjusting confidence is by increasing it as more confirming information becomes available, similar to a Bayesian style (Diaconis and Zabell 1982; Rust et al. 1999). Thus, when information from a different source that is correct 70% of the time is sequentially added to information from a source that is 80% accurate, confidence in the information should increase above 80%. Specifically, Bayes's theorem is an algebraic formulation by which prior judgments and beliefs (e.g., probabilities, confidences) are revised in light of new empirical evidence, in order to obtain posterior (updated) judgments and beliefs (Brinberg, Lynch, and Sawyer 1992), which provides a normative standard of comparison for actual judgments. In updating judgments and beliefs, Bayesian decision makers' probability judgments after the *n*th piece of information become the "prior probabilities" (proxy for base rates) for the $(n + 1)$th piece of information (e.g., McKenzie 1994). The Bayesian formulation for computing the posterior/updated probability $P(\mathrm{H1}|D)$ is

$$P(\mathrm{H1}|D) = \frac{P(\mathrm{H1})P(D|\mathrm{H1})}{P(\mathrm{H1})P(D|\mathrm{H1}) + P(\mathrm{H2})P(D|\mathrm{H2})}, \quad (1)$$

where H1 and H2 refer to mutually exclusive hypotheses, D refers to the new sequential data, and $P(D|\mathrm{H1})$ is the probability of observing the new data, given that H1 is true.

Consider the retirement investment scenario described earlier as an illustrative example. Hypothesis H1 posits that "Fidelity will perform better than TIAA-CREF," and hypothesis H2 that "TIAA-CREF will perform better than Fidelity." Suppose that, after seeing the first piece of information, the investor is 80% confident that Fidelity will perform better than TIAA-CREF—that is, $P(\mathrm{H1}) = .80$ and $P(\mathrm{H2}) = .20$. Later, the investor receives new data (i.e., D in the above equation) that another agent, who has been correct 70% of the time, also ranks Fidelity as performing better than TIAA-CREF; that is, $P(D|\mathrm{H1}) = .70$ and $P(D|\mathrm{H2}) = .30$. Given this second piece of data, a rational decision maker should increase her confidence that Fidelity will perform better than TIAA-CREF; that is, the updated confidence judgment, $P(\mathrm{H1}|D)$, should be greater than the initial confidence judgment of 80%, since the second agent's accuracy rate is better than chance (i.e., above .50 or 50%). The normative posterior judgment predicted by Bayesian theory, assuming that the two sources are independent, can be calculated from equation 1 as $(.80 \times .70)/(.80 \times .70 + .20 \times .30) = .90$. Hence, according to the Bayesian model, the investor should be 90% confident that Fidelity would perform better than TIAA-CREF.

Prior studies that examine the extent to which consumer belief updating is consistent with the normative Bayesian model offer mixed results. Some studies (e.g., Bar-Hillel 1980; Kahneman and Tversky 1972) concluded that people are "not Bayesian at all" (Kahneman and Tversky 1972, 450), while another set of studies suggested that people update directionally consistent with, but not as far as, Bayesian predictions (Edwards 1968; Rust et al. 1999). Finally, a third line of research (e.g., Cosmides and Tooby 1996; Gigerenzer and Hoffrage 1995) proposed that individuals update their beliefs in a Bayesian fashion under some conditions. In this research, we attempt to reconcile the apparent inconsistency in prior studies by examining how the data format (frequency vs. percentage) and the related issues of consumers' ability and motivation to process the data affect the extent to which consumers update their confidences in relative judgments when processing sequential data.

Updating by Averaging

In contrast to the Bayesian model, averaging is a different algorithm through which individuals can revise their judgments (Birnbaum and Mellers 1983; Johar, Jedidi, and Jacoby 1997; Shanteau 1975). When facing new information, individuals can revise their judgments by simply averaging the data (Lopes 1985; Shanteau 1975). Hence, using an averaging model with all data equally weighted, the updated probability judgment (P_{U}) can be represented as

$$P_{\mathrm{U}} = \frac{\sum_{i=1}^{n} D_i}{n}, \quad (2)$$

where D_i is each piece of data (or the belief associated with each piece of data). The averaging model is generally considered to be an effective updating strategy given its simplicity and ease of computation (Anderson 1981; Hogarth and Einhorn 1992); hence, it is often employed by decision makers (Dougherty and Shanteau 1999). Applying equation 2 to the previous investing example, the updated confidence judgment based on averaging would be $(80 + 70)/2 = 75\%$. As can be seen from this illustrative example, relative to the Bayesian model, the averaging model tends to produce more conservative updated judgments (McKenzie 1994). In fact, as long as sequential data are greater than .50, averaging will produce lower values than Bayesian predictions; moreover, for counterbalanced order of presentation, averaging will lead to no updating (i.e., changes) of judgments between the sequential data points. (Note that across all our studies, we only use scenarios with values above .50 and counterbalanced order of presentation.)

Data Format: Percentage versus Frequency

Percentage and frequency are two common formats for presenting data and are equivalent in terms of information conveyed regarding event likelihood. Practitioners have used the two data formats interchangeably to present source accuracy data (see also Chatterjee et al. [2000], Chen and Rao [2007], DelVecchio, Krishnan, and Smith [2007], and Heath, Chatterjee, and France [1995] for additional contexts of percentage vs. frequency data being used).

Studies that found that individuals are likely to use the averaging rule in belief updating have incidentally used stimuli data in percentage format—for example, witness 1 was correct 80% of the time, and witness 2 was correct 70% of the time (Hogarth and Einhorn 1992; Lopes 1985; Shanteau 1975). In contrast, another stream of research suggested that frequency format facilitates the use of base rate information and the application of the Bayesian theorem (Gigerenzer and Hoffrage 1995, 1999). Specifically, Gigerenzer and Hoffrage (1995) examined cases where individuals were given base rate information and were then asked to make a probability judgment based on a single piece of additional information (also known as "one-shot" task; McKenzie 1994). Early work in this research stream used percentage or probability formats (e.g., "If a woman has breast cancer, the probability is 80% that she will get a true positive mammography"). One major finding has been that individuals tend to ignore base rate information in making probability judgments (McKenzie 1994). However, Gigerenzer and Hoffrage (1995) demonstrated that individuals were more likely to utilize and incorporate base rate information and make likelihood judgments in line with Bayesian principles when the problems were presented in frequency format (e.g., "8 out of 10 women with breast cancer will get a positive mammography") rather than in probability/percentage format (e.g., "if a woman has breast cancer, the probability is 80% that she will get a positive mammography"). Gigerenzer and Hoffrage (1995) concluded that a frequency format could enhance Bayesian reasoning (however, see Lewis and Keren [1999] and Mellers and McGraw [1999] for alternative views).

Processing of Percentage versus Frequency Data

Individuals' knowledge about their own cognitive processes and/or abilities and the consequent regulation of their cognitive resources influence their choice of processing strategy (Garafalo and Lester 1985; Metcalfe 2009). As a result, the perceived ease or difficulty of processing information is likely to affect processing strategy. Prior studies have examined the effects of ease or difficulty of processing (i.e., processing fluency) in the context of varied issues such as individual judgments of liking, preference, attitude, and persuasion (Schwarz 2004). In contrast, no study has examined how the difficulty in processing sequential data might influence individuals' updating of confidences in relative judgments. We propose that individuals' perceived difficulty of processing sequential data influences their choice of algorithm for processing the data, which in turn influences their confidence-judgment updating.

In essence, individuals are cognitive misers who are likely to abandon a mathematical algorithm if they experience difficulty in employing it (Shugan 1980). Research in the domain of mathematical computations has noted that in processing numerical data, one needs to determine an applicable algorithm (Kantowski 1977) and also have the motivation and ability to execute that algorithm (Dehaene 1997). Lacking motivation or ability (e.g., due to perceived difficulty) to execute a particular mathematical algorithm, a consumer will abandon that algorithm and resort to heuristics or shortcuts.

When evaluating a sequence of percentage data, consumers are likely to use an algorithm similar to averaging, especially given the relative ease of doing so (Dougherty and Shanteau 1999). In contrast, computing the average of a sequence of data in frequency format is relatively much more difficult, since sequential frequency data are a series of fractions and hence participants have to undertake multiple computational steps—they first have to convert the data into percentages and then average it (or combine the numerators and the denominators and then compute the ratio; Dehaene 1992). Hence, averaging sequential frequency data involves multiple arithmetic processes, which is more cognitively demanding (Dehaene 1992; Peterson and Aller 1971). Consumers, being cognitive misers, would have little motivation to allocate the cognitive resources to compute averages and would tend to resort to heuristics instead (e.g., Payne 1976). Moreover, from the perspective of an adaptive decision maker, an effective processing strategy would be to rely on heuristics instead of systematic processing, especially when the decision maker is not able and/or motivated in making the systematic processing (Newell and Simon 1972; Payne, Bettman, and Johnson 1988, 1993); this is because heuristic processing will require fewer operations than systematic processing (e.g., in the form of computing the averages of sequential frequency data) and will hence be more effective.

When facing processing difficulty, one heuristic for integrating multiple pieces of frequency data would be to have confirmatory belief updating; that is, a consumer can form a hypothesis based on the first piece of datum and then use subsequent sequential data to directionally confirm or disconfirm that hypothesis (Sanbonmatsu, Posavac, and Stasney 1997; Sanbonmatsu et al. 1998). For sequential frequency data, the first piece of information becomes the focal hypothesis (Posavac et al. 2004; Sanbonmatsu et al. 1997, 1998), and additional pieces of information help directionally confirm or disconfirm this hypothesis (e.g., Klayman and Ha 1987; Nickerson 1998; Poletiek and Berndsen 2000). When subsequent data are in the same direction as the initial piece of information, a consumer's initial hypothesis is confirmed, and hence there will be an upward revision of the consumer's confidence in the hypothesis. For instance, in the retirement fund illustrative example, if participants are told that one agent (who has been correct 128 out of 177 times in the past) has ranked Fidelity as likely to perform better than TIAA-CREF, most participants will be unable and/or unmotivated to determine the exact equivalent percentage values. Instead, they will form a hypothesis that "Fidelity will perform better than TIAA-CREF" with a degree of confidence based on the given data (i.e., "it *seems* there is approximately a 70% chance that Fidelity will perform better than TIAA-CREF"). When a second piece of information is presented from another agent, who also ranked Fidelity as performing better than TIAA-CREF, the hypothesis will be confirmed and confidence in the judgment will be strengthened (i.e., go up from 70%). That is, consumers are likely to increase their confidence in the relative judgment that "Fidelity will perform better than TIAA-CREF."

Interestingly, as can be seen from equation 1, the Bayesian model also prescribes a similar hypothesis-testing algorithm, since as long as the new sequential data is greater than .50, the posterior/updated judgment will increase. Thus, we expect that when sequential data are difficult to process mathematically (e.g., frequency data), consumers' updated confidences in relative judgments would be in a similar direction as the Bayesian model. In contrast, when the data are presented in a relatively easy-to-compute format (e.g., percentages), we expect consumers will revise their judgments by averaging and hence will have lower updated judgments. Formally stated, when sequential counterbalanced data have accuracy rates above .50:

H1:

a) For sequential frequency data, consumers will update their confidences in relative judgments in an upward direction from the first data point to the second.

b) For sequential percentage data, consumers will not update their confidences in relative judgments from the first data point to the second.

H2:

a) For sequential frequency data, consumers' updated confidences in relative judgments will tend toward the Bayesian model and will be higher than the averaging model.

b) For sequential percentage data, consumers' updated confidences in relative judgments will be lower than the Bayesian model but similar to the averaging model.

STUDY 1: PERCENTAGE VERSUS FREQUENCY DATA INVOLVING FINANCIAL PRODUCTS

Design and Participants

Hypotheses 1 and 2 were tested in study 1 with the help of a 2 (data format: percentage vs. frequency) × 2 (repeated measures of confidence in relative judgment) mixed factor design experiment. The first factor was manipulated between subjects, and the second factor was within subjects or repeated measures—participants stated confidence in relative judgment after they saw each of the two pieces of sequential data. Seventy-five university students participated in exchange for course credit (average age 21 years, 72% females).

Procedure and Dependent Variables

When making relative judgments, consumers frequently consider two-option scenarios. Even when choosing from a larger consideration set, consumers often undertake ordinal comparisons between two options at a time for ease of evaluation (Shugan 1980). Hence, not surprisingly, prior research on ordinal rankings has focused on two-option scenarios (e.g., Fox and Levav 2000). Consistent with this, in our experiments, we also examine consumer updating of confidence in relative judgment in the context of two options.

Stimuli presentation and data collection were done through a computer. Participants were asked to read sequential reports from two business publications (*Fortune* and *Forbes*) regarding the likely performances of two different mutual funds—Advance Capital Equity Fund (ACE) and BB&T Equity Fund. Participants were also given information regarding *Fortune*'s and *Forbes*'s past accuracy rates in ranking mutual fund performance. The two pieces of data were sequentially introduced (e.g., Gürhan-Canli 2003). Consistent with actual managerial practices (e.g., Simons 1998; Vinod 2004), participants were given information about the ordinal (relative) rankings of the mutual funds instead of cardinal ratings. In the given scenario, both *Fortune* and *Forbes* ranked the BB&T fund better than the ACE fund in terms of future potential returns and growth. The past accuracy rates of *Fortune* and *Forbes* were given as 70.5% (out of 23 times) and 69.5% (out of 17 times) in the percentage data condition and as 12 (out of 17 times) and 16 (out of 23 times) in the frequency data condition. Although there was low variance between the accuracy rates of the two pieces of data (70.5% and 69.5%), the data were counterbalanced to avoid any potential order effects (Hogarth and Einhorn 1992). After seeing the first datum, participants responded to a set of questions, two of which measured participant confidence in the relative judgment that BB&T would fare better than ACE in the future. After this, participants saw the second datum and again responded to a set of questions, including the ones measuring the updated confidence in relative judgment regarding BB&T being better than ACE.

Two key dependent variables were included in this study. First, participant confidence in relative (ordinal) judgment was operationalized as strength of belief or confidence with the probability statement about which one of two items is better and was measured by taking the mean of two questions: "How confident are you that BB&T Equity Fund would perform better than Advance Capital Equity Fund in terms of future potential returns and growth? (0 = No Confidence; 100 = 100% Confidence)" and " In your opinion, based on the given information, on a scale of 0–100, what is the probability that BB&T Equity Fund would fare better than Advance Capital Equity Fund in the future?" ($r = .74$ for initial judgment measures and $r = .72$ for updated judgment measures). Second, in order to examine whether participants' allocation of money to the two mutual funds reflected the confidence they had in their relative judgments, participants were given a hypothetical scenario where they had $1,000 to invest and had the option of investing in either one or both of these funds in whatever proportion they chose.

Results

Main Tests. A 2 (percentage vs. frequency) × 2 (repeated confidence in relative judgment measures) mixed ANOVA revealed a significant interaction effect ($F(1, 73) = 29.84$, $p < .01$). Consistent with hypothesis 1, when data were presented in frequency format, participants updated their confidences in relative judgments in an upward direction, from the first data point to the second ($M_{\text{initial}} = 64.86$ vs. $M_{\text{update}} = 78.84$; $F(1, 73) = 70.97$, $p < .01$). In contrast, and as expected, when data were presented in percentage format, participants did not significantly update (i.e., change) their confidences in relative judgments from the first data point to the second ($M_{\text{initial}} = 65.56$ vs. $M_{\text{update}} = 66.64$; $F(1, 73) = .41$, $p = .53$).

Consistent with the predictions made by hypothesis 2, when data were presented in frequency format, consumers' updated confidences in relative judgments tended toward those made by the normative Bayesian model (as calculated from eq. 1; $M_{\text{update}} = 78.84$ vs. $M_{\text{Bayes}} = 80.76$; $t(37) = 1.49$, $p = .14$) but were higher than the averaging model ($M_{\text{update}} = 78.84$ vs. $M_{\text{avg}} = 70.0$; $t(37) = 8.25$, $p < .01$). In contrast, when data were presented in percentage format, consumers' updated confidence judgments were lower than those made by the Bayesian model ($M_{\text{update}} = 66.64$ vs. $M_{\text{Bayes}} = 80.46$; $t(36) = 7.84$, $p < .01$) but were similar to the averaging model ($M_{\text{update}} = 66.64$ vs. $M_{\text{avg}} = 70.0$; $t(36) = 1.25$, $p = .22$).

Participants' allocation of money (out of $1,000) to each of the two mutual funds reflected their updated confidences in relative judgments. Specifically, as expected, when sequential product data were presented in frequency format, participants' allocation of money (out of $1,000) for the higher-ranked mutual fund (BB&T) tended toward Bayesian predictions ($M_{\text{BB\&T}} = 779.73$ vs. $M_{\text{Bayes}} = 807.62$; $t(36) = 1.39$, $p = .17$) and were higher than the averaging model ($M_{\text{BB\&T}} = 779.73$ vs. $M_{\text{avg}} = 700$; $t(36) = 5.58$, $p < .01$), while for percentage data, participants' allocation of money for the higher-ranked fund was lower than Bayesian predictions ($M_{\text{BB\&T}} = 718.38$ vs. $M_{\text{Bayes}} = 804.63$; $t(36) = 4.25$, $p < .01$); instead, for percentage data, participants' allocations of money to the funds were consistent with the averaging model ($M_{\text{BB\&T}} = 718.38$ vs. $M_{\text{avg}} = 700$; $t(36) = 1.01$, $p = .32$).

Process Results. We wanted to check whether participants perceived sequential frequency data to be computationally more difficult than sequential percentage data. Participants were asked how difficult it was to mathematically process the given sequential data (1 = not at all difficult, 7 = extremely difficult). As expected, participants indicated higher perceived difficulty for the frequency than the percentage format ($M_{\text{freq}} = 4.82$ vs. $M_{\text{percent}} = 3.11$; $t(73) = 7.38$, $p < .01$). To further gauge the underlying process, participants were also asked to indicate the extent to which they averaged the two given pieces of information from *Fortune* and *Forbes* (1 = did not average at all, 7 = averaged to a great extent). As expected, participants averaged to a greater extent for percentage than for frequency data ($M_{\text{percent}} = 4.68$ vs. $M_{\text{freq}} = 3.63$; $t(73) = 3.63$, $p < .01$). Finally, to examine the extent to which participants used the second piece of data to confirm their hypothesis related to the first piece of data, they were asked to indicate the extent to which the second piece of information confirmed the first piece of information (1 = very little extent, 7 = great extent). As expected, participants perceived the second piece of information to be more confirmatory of the first piece when the data were presented in frequency than in percentage format ($M_{\text{freq}} = 4.71$ vs. $M_{\text{percent}} = 3.97$, $t(73) = 3.51$, $p < .01$). Hence, these three measures provided convergent evidence supporting our theorization of the underlying process through which consumers differentially update their confidence in relative judgment for percentage versus frequency data.

Ruling Out an Alternative Explanation. We attempted to examine whether memory played a role in terms of differential degrees of recalls of the sequential data. Toward the end of the survey, participants were asked to recall the sequential data that they read earlier. The two pieces of sequential data (first and second) were recalled by an equal proportion of participants (.89 vs. .91; $z = .33$, $p = .74$), and this pattern of results was observed for percentage data (.89 vs. .92; $z = .37$, $p = .71$) as well as for frequency data (.89 vs. .89; $z = 0$, $p = 1$).

Discussion

Study 1 showed that consumers tend to update their confidences in relative judgments to a greater extent for sequential frequency data than for percentage data. As a result, consumers' updated confidence judgments were relatively closer to the normative Bayesian model when data were presented in frequency, rather than percentage, format. Instead, for percentage data, updated confidence judgments were consistent with the averaging model, while for frequency data, updated confidence judgments were higher than the averaging model. The results also show that investors are likely to allocate money to different investment opportunities in a manner consistent with their updated confidence judgments.

Though the process measures in study 1 provide empirical evidence supporting our theorizing that it is the processing difficulty of frequency data that causes individuals to abandon an algorithm consistent with the averaging model, we did not directly manipulate the computation difficulty level of frequency data. In study 2, we extend the findings of study 1 by directly manipulating the computation difficulty level of frequency data and also replicate the key findings of study 1 with a different product category (car).

STUDY 2: EASY- VERSUS DIFFICULT-TO-COMPUTE FREQUENCY DATA

Study 2 examined three data format conditions: percentage, relatively easy-to-compute frequency, and relatively difficult-to-compute frequency. Frequency data with denominators of 100, 200, and 50 are relatively easy to convert to equivalent percentage values. Based on our theorization, confidence in relative judgments would be updated to a greater extent for relatively difficult-to-compute frequency data than for relatively easy-to-compute frequency data or for percentage data. That is, for sequential data in relatively easy-to-compute frequency or in percentage format, consumers are more likely to use an algorithm consistent with averaging.

Design and Subjects

Study 2 used a 3 (data format: percentage data vs. relatively easy-to-compute frequency data vs. relatively difficult-to-compute frequency data) × 2 (repeated measures of confidence in relative judgment) mixed factor design. As in study 1, the first factor was manipulated between subjects, and the second factor was within subjects or repeated measures. Participants rated their confidences in relative judgments after they saw each of the two pieces of sequential data. Sixty-nine university students participated for extra course credit (average age 22 years, 41% females).

Procedure and Dependent Variables

The procedure and key measures were similar to those used in study 1, with the key change being in the type of product. Participants were asked to read safety rankings for two cars (labeled A and B) by two agencies—the National Highway Traffic Safety Administration (NHTSA) and the Insurance Institute for Highway Safety (IIHS); both agencies ranked Car B as safer than Car A. The past accuracy rates of NHTSA and IIHS, respectively, were 70% and 71% (out of 173 times) for percentage data, 140 and 142 (out of 200 times) for easy-to-compute frequency data, and 121 and 123 (out of 173 times) for difficult-to-compute frequency data. The data were sequentially introduced, with the order counterbalanced. After each of the two pieces of sequential data, participants responded to a set of questions, two of which measured participants' confidences in their relative judgments that Car B was safer than Car A. As in study 1, participant confidence in relative judgment was measured by taking the mean of two questions related to the perceived confidence in the probability that Car B is safer than Car A ($r = .87$ for initial judgment measures and $r = .83$ for updated judgment measures).

Results

A 3 × 2 mixed ANOVA revealed an interaction effect ($F(2, 66) = 6.35$, $p < .01$). Consistent with hypothesis 1, for relatively difficult-to-compute frequency data, there was a significant updating of confidence in relative judgment from the first to the second data point (65.0 vs. 77.35; $F(1, 66) = 25.89$, $p < .01$); there was no significant updating (i.e., changes) for the easy-to-compute frequency data (66.97 vs. 70.66; $F(1, 66) = 1.90$, $p = .17$) and the percentage data (67.11 vs. 67.98; $F(1, 66) = .15$, $p = .70$). The updated confidence in relative judgment for the difficult-to-compute frequency condition was higher than those for both the easy-to-compute frequency ($t(40) = 2.78$, $p < .01$) and percentage $t(48) = 3.17$, $p < .01$) conditions, with the latter two conditions being equivalent to each other ($t(44) = .90$, $p = .37$).

Also, consistent with hypothesis 2, for relatively difficult-to-compute frequency data, participants' updated confidences in their relative judgments tended toward the Bayesian model ($M_{\text{update}} = 77.35$ vs. $M_{\text{Bayes}} = 80.09$; $t(22) = .84$, $p = .41$) and were higher than the averaging model ($M_{\text{avg}} = 70.5$; $t(22) = 3.80$, $p < .01$). In contrast, for both the percentage data and the relatively easy-to-compute frequency data, participants' updated confidence levels were significantly lower than those of the Bayesian model (for percentage data, $M_{\text{update}} = 67.98$ vs. $M_{\text{Bayes}} = 82.51$; $t(26) = 10.97$, $p < .01$; for relatively easy-to-compute frequency data, $M_{\text{update}} = 70.66$ vs. $M_{\text{Bayes}} = 82.46$; $t(18) = 7.08$, $p < .01$). Instead, updated confidences in relative judgments were similar to the averaging model ($M_{\text{avg}} = 70.5$) for percentage data ($t(26) = 1.12$, $p = .27$), as well as for relatively easy-to-compute frequency data ($t(18) = .11$, $p = .92$).

Discussion

Study 2 replicated and extended the findings of study 1 by demonstrating that for difficult-to-compute frequency data, consumer confidence in relative judgment is significantly updated, tending toward the Bayesian model, and higher than the averaging model. For percentage data, and for relatively easy-to-compute frequency data, however, updated confidence judgments were similar to an averaging model and hence much lower than Bayesian predictions. Next, study 3 examines the moderating effects of consumer processing motivation and, in the process, attempts to provide further support for our theorizing.

STUDY 3: EFFECTS OF PROCESSING MOTIVATION

We theorized that for sequential data in percentages, consumers are likely to revise their confidences in relative judgments in a manner similar to averaging. For frequency data, however, most consumers would not have the motivation and ability (due to processing difficulty) to apply a mathematical algorithm in the form of averaging the data; instead there will be confirmatory belief updating, similar to a Bayesian model. Processing motivation should not have any impact while evaluating sequential percentage data, given the relative ease of using an averaging algorithm for such data. However, for sequential frequency data, when processing motivation is enhanced, consumers are more likely to undertake systematic mathematical computations. That is, under high processing motivation, for sequential frequency data, consumer judgment updating would tend toward an averaging algorithm and would hence be similar to what happens for sequential percentage data. Therefore, for sequential frequency data, the effects of hypotheses 1 and 2 would hold only when consumers have low processing motivation, with the effects getting reduced under high processing motivation. Formally stated:

H3:

a) For sequential frequency data, consumers will update their confidences in relative judgments in an upward direction from the first data point to the second, when processing motivation is low, with the effects getting attenuated when processing motivation is high.

b) For sequential percentage data, consumers will not update their confidences in relative judgments from the first data point to the second, irrespective of processing motivation level.

H4:

a) For sequential frequency data, consumers' updated confidences in relative judgments will tend toward the Bayesian model when processing motivation is low but will be lower than the Bayesian model and similar to the averaging model when processing motivation is high.

b) For sequential percentage data, consumers' updated confidences in relative judgments will be lower than the Bayesian model and similar to the averaging model, irrespective of processing motivation level.

Method

A 2 (data format: percentage vs. frequency) × 2 (processing motivation: low vs. high) × 2 (repeated measures of confidence in relative judgment) mixed factor design experiment was used to test hypotheses 3 and 4. The first two factors were manipulated between subjects, and the third factor was within subjects or repeated measures. Ninety-nine university students participated in this study for extra course credit (average age 22 years, 57% females).

Stimuli presentation and data collection were done through a computer, which also unobtrusively recorded participants' response latencies for key variables. The procedure and product (mutual fund) were similar to those used in study 1. That is, participants were asked to read sequential reports from two sources (*Zacks* and *Morningstar*) regarding the likely performances of two different mutual funds—ACE and BB&T. Participants were also given information regarding past accuracy rates of *Zacks* and *Morningstar* for similar types of two-fund rankings. Both *Zacks* and *Morningstar* ranked the BB&T fund better than the ACE fund in terms of future potential returns and growth. In the frequency condition, participants were told that when two mutual funds were compared, *Zacks* (and *Morningstar*) made the correct ranking prediction 51 times out of 73 times (first sequential datum) and 54 times out of 77 times (second sequential datum) in the past. In the percentage condition, the equivalent percentage values (i.e., 70%) were given. For both the frequency and percentage conditions, the data were counterbalanced to control for potential order effects. Participants reported their confidence in relative judgment after each sequential datum, which was measured in a similar manner as in study 1 ($r = .62$ for initial confidence judgment measures and $r = .72$ for updated confidence judgment measures).

Participant processing motivation was manipulated in a manner similar to that adopted in prior studies (e.g., Shiv, Edell-Britton, and Payne 2004). That is, in the high processing motivation conditions, participants were told that their opinions are extremely important and will be analyzed individually by the researcher; hence, their individual opinions will have tremendous implication for this research. In the low processing motivation conditions, participants were told that their opinions will be combined with those of other participants and will be analyzed at the aggregate level by the researcher; hence, individual opinions will not have much of an implication for this research.

Results

Manipulation Checks. To examine the successful manipulation of processing motivation, participants indicated on three items (e.g., Shiv et al. 2004), anchored by disagree (1) and agree (7), the extent to which they found the given scenario and accompanying information interesting, involving, and personally relevant ($\alpha = .81$). As expected, participants' ratings were higher in the high (vs. low) motivation condition ($M = 4.26$ vs. 3.10; $F(1, 97) = 52.20$, $p < .01$).

Main Tests. The results of a 2 (data format) × 2 (motivation) × 2 (confidence in relative judgment) mixed

ANOVA showed an interaction effect ($F(1, 95) = 6.89$, $p < .01$). Consistent with hypothesis 3a, when data were presented in frequency format, participants updated their confidences in relative judgments in an upward direction, from the first data point to the second ($M_{\text{initial}} = 66.12$ vs. $M_{\text{update}} = 78.75$; $F(1, 95) = 38.01$, $p < .01$) only when processing motivation was low, but not under high processing motivation ($M_{\text{initial}} = 68.12$ vs. $M_{\text{update}} = 68.48$; $F(1, 95) = .03$, $p = .86$). In contrast, and consistent with hypothesis 3b, when data were presented in percentage format, there were no statistically significant changes in confidence judgments from the first data point to the second for either low ($M_{\text{initial}} = 67.98$ vs. $M_{\text{update}} = 71.17$; $F(1, 95) = 2.15$, $p = .15$) or high ($M_{\text{initial}} = 67.70$ vs. $M_{\text{update}} = 69.66$; $F(1, 95) = .88$, $p = .35$) motivation.

Consistent with the predictions made by hypothesis 4a, when data were presented in frequency format, updated confidence in relative judgment tended toward the Bayesian model ($M_{\text{update}} = 78.75$ vs. $M_{\text{Bayes}} = 81.51$; $t(25) = 1.62$, $p = .12$) and was higher than the averaging model ($M_{\text{avg}} = 70.0$; $t(25) = 6.86$, $p < .01$) only when processing motivation was low; when processing motivation was high, updated confidence in relative judgment was lower than the Bayesian prediction ($M_{\text{update}} = 68.48$ vs. $M_{\text{Bayes}} = 82.75$; $t(24) = 6.32$, $p < .01$) but similar to the averaging model ($t(24) = .53$, $p = .60$). In contrast, when data were presented in percentage format, updated confidence in relative judgment was lower than the Bayesian model, irrespective of whether processing motivation was low ($M_{\text{update}} = 71.17$ vs. $M_{\text{Bayes}} = 82.73$; $t(22) = 5.12$, $p < .01$) or high ($M_{\text{update}} = 69.66$ vs. $M_{\text{Bayes}} = 82.43$; $t(24) = 8.85$, $p < .01$); instead, updated confidence in relative judgment was similar to the averaging model (for low motivation, $t(22) = .40$, $p = .69$; for high motivation, $t(24) = .17$, $p = .86$).

Process Results. We theorized that when processing motivation is high, consumers are more likely to use mathematical algorithms for frequency data. Hence, consumers should take more time to make confidence judgments after seeing the second sequential frequency data when processing motivation is high (vs. low). Consistent with this theorizing, for frequency data, participants' response latencies were higher when processing motivation was high versus low ($M_{\text{high}} = 21.41$ seconds vs. $M_{\text{low}} = 11.11$ seconds; $t(49) = 3.23$, $p < .01$).

Discussion

As hypothesized, study 3 showed some paradoxical results. While prior research has shown that higher processing motivation enhances judgment accuracy (Chen and Rao 2007), we find that higher processing motivation may reduce judgment accuracy in the case of sequential frequency data. This occurs because while higher motivation induces consumers to undertake mathematical computations, the computations are similar to those of an averaging algorithm and hence lower than the predictions of the normative Bayesian model. The results of study 3 also provided additional empirical evidence that processing motivation influences consumer choice of algorithm for evaluating sequential frequency data. Next, study 4 examines the moderating effects of highlighting the independence of sequential data.

STUDY 4: EFFECTS OF HIGHLIGHTING INDEPENDENCE OF SEQUENTIAL DATA

Research has shown that frequency data tend to create more vivid mental imagery than percentage data do (Peters et al. 2006; Slovic, Monahan, and MacGregor 2000). As a result, for sequential frequency (vs. percentage) data, consumers might see each piece of sequential data as more distinctive and independent. This can be a potential alternative explanation for the lack of sufficient judgment updating with sequential percentage data, compared to frequency data. Thus, in study 4, we examined the effects of explicitly highlighting the independence of each piece of data.

Highlighting data independence should not make any difference to the degree of judgment updating in the case of sequential frequency data, since consumers are already likely to perceive a distinction between the two pieces of data. However, for percentage data, highlighting data independence, compared with the absence of such highlighting, is more likely to lead to enhanced judgments. Formally, we predict:

H5:

a) For sequential frequency data, consumers will update their confidences in relative judgments from the first data point to the second irrespective of whether data independence is highlighted.

b) For sequential percentage data, consumers will update their confidences in relative judgments from the first data point to the second when data independence is highlighted, but there will be no judgment updating when data independence is not highlighted.

H6:

a) For sequential frequency data, consumers will have similar levels of updated confidences in relative judgments when data independence is highlighted versus when it is not.

b) For sequential percentage data, consumers will have higher updated confidences in relative judgments when data independence is highlighted (versus when it is not).

Method

A 2 (data format: percentage vs. frequency) × 2 (data independence: highlighted vs. not highlighted) × 2 (repeated measures of confidence in relative judgment) mixed factorial design experiment was used to test hypotheses 5 and 6. The first two factors were manipulated between subjects, and the third factor was within subjects or repeated measures. Ninety-two university students participated in this study for extra course credit (average age 22 years, 52% females).

The procedure was similar to those used in our other studies, with the key differences being in the product used (i.e., laptop computer) and with moderate variance in the stimuli data in terms of sources' past accuracies. Participants read sequential pieces of information from *Consumer Reports* and *CNET* regarding two laptops (A and B), whereby both sources ranked Laptop B as better than Laptop A. Moreover, participants were given past accuracy rates of these two sources/agencies as 75% (out of 177 times) and 65% (out of 183 times) in the percentage data condition and as 133 (out of 177 times) and 119 (out of 183 times) in the frequency data condition. The data order was counterbalanced, as in studies 1–3.

To manipulate data independence, participants were reminded twice that "*Consumer Reports* and *CNET* are completely independent of each other. That is, they are two different and independent sources, with no business or any other connections between them." To check whether the manipulation of data independence was effective, toward the end of the survey, participants were asked three questions: how distinctively independent the given sources are (1 = low, 7 = high), how prominent it is in their mind that the product rankings are from separate sources (1 = low, 7 = high), and the degree to which they agreed that the sources are distinctively independent from each other (1 = strongly disagree, 7 = strongly agree; $\alpha = .74$). As in study 1, the key dependent variable of confidence in relative judgment was measured by taking the mean of two questions related to the perceived confidence in the probability that Laptop B is better than Laptop A ($r = .75$ for initial judgment measures and $r = .90$ for updated judgment measures).

Results

Manipulation Checks. Consistent with the manipulations, participants perceived a higher degree of independence of sequential data when the data points' independence was highlighted ($M_{\text{highlight}} = 4.53$ vs. $M_{\text{nonhighlight}} = 3.83$; $F(1, 90) = 6.99$, $p < .01$). Also, as expected, the manipulation did not influence perceptions in the case of frequency data ($M_{\text{highlight}} = 4.56$ vs. $M_{\text{nonhighlight}} = 4.14$; $F(1, 88) = 1.23$, $p = .27$) but did boost the perceived independence of sequential percentage data ($M_{\text{highlight}} = 4.50$ vs. $M_{\text{nonhighlight}} = 3.53$; $F(1, 88) = 6.79$, $p < .05$).

Main Tests. The results of a 2 (data format) × 2 (data-independence highlighting) × 2 (confidence in relative judgment) mixed ANOVA showed a marginally significant three-way interaction effect ($F(1, 88) = 2.97$, $p < .09$). Replicating the findings of studies 1–3, participants' updated confidences in relative judgments tended toward the Bayesian model when the data were in frequency format irrespective of data-independence highlighting (for highlighted data: $M_{\text{update}} = 84.28$ vs. $M_{\text{Bayes}} = 86.04$; $t(22) = 1.33$, $p = .19$; for nonhighlighted data: $M_{\text{update}} = 81.39$ vs. $M_{\text{Bayes}} = 84.04$; $t(22) = 1.49$, $p = .15$), and these were higher than the averaging model ($p < .01$). In contrast, for percentage data, updated confidence judgments were lower than the Bayesian model irrespective of data independence highlighting (for highlighted data: $M_{\text{update}} = 73.52$ vs. $M_{\text{Bayes}} = 83.33$; $t(21) = 7.51$, $p < .01$; for nonhighlighted data: $M_{\text{update}} = 66.77$ vs. $M_{\text{Bayes}} = 81.38$; $t(23) = 8.76$, $p < .01$); for highlighted percentage data, updated judgments were higher than the averaging model ($t(21) = 2.51$ $p < .05$), while for nonhighlighted percentage data, updated judgments were similar to the averaging model ($t(23) = 1.53$, $p = .14$).

Consistent with hypothesis 5, for frequency data, participants updated their confidence judgments regardless of whether data independence was highlighted ($M_{\text{initial}} = 74.89$ vs. $M_{\text{update}} = 84.28$; $F(1, 88) = 21.74$, $p < .01$) or not ($M_{\text{initial}} = 70.07$ vs. $M_{\text{update}} = 81.39$; $F(1, 88) = 31.61$, $p < .01$). In contrast, for percentage data, participants did not update their confidence judgments when data independence was not highlighted ($M_{\text{initial}} = 66.19$ vs. $M_{\text{update}} = 66.77$; $F(1, 88) = .09$, $p = .77$) but did update them when data independence was highlighted ($M_{\text{initial}} = 67.93$ vs. $M_{\text{update}} = 73.52$; $F(1, 88) = 7.37$, $p < .01$). In other words, for percentage data, participants' confidences in relative judgments increased only when data independence was highlighted.

Consistent with hypothesis 6, for frequency data, participants had similar updated confidence judgments irrespective of whether data independence was highlighted ($M_{\text{highlight}} = 84.28$ vs. $M_{\text{nonhighlight}} = 81.39$; $t(44) = 1.38$, $p = .17$), while for percentage data, updated confidence judgments were higher when data independence was highlighted versus when it was not ($M_{\text{highlight}} = 73.52$ vs. $M_{\text{nonhighlight}} = 66.77$; $t(44) = 2.67$, $p < .05$). Moreover, frequency data led to higher updated confidence judgments than percentage data irrespective of whether the data were highlighted ($t(43) = 4.96$, $p < .01$) or not ($t(45) = 5.86$, $p < .01$).

Discussion

Study 4 replicated the key results obtained in studies 1–3, using moderate variance in the sources' past accuracies. Also, for percentage data, consumers updated their confidences in relative judgments when data independence was highlighted, although still more conservatively than the Bayesian prediction. In contrast, for frequency data, consumers updated their confidences in relative judgments to a similar extent irrespective of whether data independence was highlighted, with updated judgments tending toward the Bayesian model.

GENERAL DISCUSSION

Summary and Conclusions

Four experiments show that when consumers encounter sequential product ranking and source accuracy data, the updating of confidence in relative judgment is influenced by the data format (percentage vs. frequency) in which source accuracy data is presented. Specifically, consumers' confidence judgments are updated more and tend toward the normative Bayesian model when sequential source accuracy data are given in frequency (rather than percentage) format. In contrast, when the data are in percentage format, consumers' updated confidence judgments are consistent with an averaging model. It might be noted that while our empirical results showed that for frequency data updated confidence judgments were statistically equivalent to the Bayesian model, our conceptual framework does not make such a strong claim; instead, we are proposing that for frequency data, updated confidence judgments will be greater than the averaging model and will tend toward the Bayesian model, while for percentage data, updated confidence judgments will be lower than the Bayesian model.

The findings of our research have relevance for how one might model an updating process. There is a significant body of theoretical and empirical work in the marketing science literature that assumes normative Bayesian updating by consumers (e.g., Erdem et al. 2005; Kopalle and Lehmann 2001). However, if consumers do not update in a Bayesian fashion, several normative conclusions from this literature may need rethinking. The results of our studies show that under certain conditions, consumers do not update in a manner consistent with the Bayesian model, which has definite relevance for how one might model an updating process.

Our results are related to the concept of "number sense" proposed by Dehaene and coauthors (e.g., Dehaene 1997, 2001; Izard and Dehaene 2008). They found that when individuals are unable to mathematically compute or process numerical data, they form a "number sense" that represents the data analogically and approximately (Izard and Dehaene 2008). Our findings are also consistent with the notion of elementary information processing strategies that an adaptive decision maker would choose, trading off judgment accuracy with effort while making decisions (Newell and Simon 1972; Payne et al. 1988, 1993). We theorize that when consumers are able to easily process the data mathematically (e.g., in percentage format), they use an algorithm similar to an averaging model. However, when it is more difficult to mathematically process the data (e.g., in frequency format), consumers form an approximate estimate (or hypothesis) for their relative judgments based on the first piece of data. When they see confirming evidence of the approximation (or hypothesis), they strengthen their belief or confidence in an upward direction, consistent with a Bayesian model. The findings of our research are interesting since we demonstrate scenarios in which a simplifying heuristic appears to improve (confidence) judgments. Such a finding also has implications for error-effort trade-off.

Prior research on third-party sources has focused on issues such as the extent to which consumers seek others' opinions (Swaminathan 2003), how consumers choose among multiple opinion sources based on the perceived diagnosticity of the source (Gershoff et al. 2001), and how perceived diagnosticity of information source is affected by consumer characteristics such as aspiration level (West and Broniarczyk 1998), agent characteristics such as opinion extremity (Gershoff, Mukherjee, Mukhopadhyay 2003), and decision tasks such as seeking brand recommendations versus seeking brand evaluations (Gershoff et al. 2001). However, to the best of our knowledge, no prior study has examined how consumers update confidences in relative judgments as they obtain sequential information on product rankings and the sources' past accuracy levels when the data are presented in percentage versus frequency formats.

Finally, our research also shows that for sequential frequency data, increasing processing motivation can lead to updated confidence judgments that are normatively less accurate. That is, while prior research has shown that higher processing motivation enhances judgment accuracy (Chen and Rao 2007), we find that higher processing motivation reduces judgment accuracy in the case of sequential frequency data. This occurs because the higher processing motivation induces consumers to average the separate accuracy rates, leading to updated judgments that are lower than the normative Bayesian model.

Managerial and Regulatory Implications

The findings have intriguing implications. For instance, when trying to discourage smoking behavior and highlighting the risks of smoking, regulators might want to use frequency format for sequential data presentation, since consumer judgment updating seems to be greater (and tending toward the normative Bayesian model) for such a data format. This is especially relevant since young people underestimate the severity of risks associated with smoking and their personal vulnerability to such potential risks (Pechmann and Shih 1999). While regulators (e.g., the U.S. surgeon general's office) and nonprofit organizations (e.g., the American Cancer Society) use both percentage and frequency formats for presenting statistical data, they do seem to use frequency format to a greater extent. For instance, as mentioned earlier, the surgeon general's 2004 report presents sequential data (across different pages) regarding cancer caused by smoking, mostly in frequency format, such as "An estimated 171,900 new cases [of lung cancer] and an estimated 157,200 deaths" and "An estimated 22,400 new cases [of stomach cancer] and 12,100 deaths." Our research indicates this to be an appropriate format for presenting this data. In contrast, when presenting adverse reaction data for prescription drugs, firms often tend to use percentage formats. For example, in package inserts and printed advertisements, the prescription drug Prozac mostly uses percentages to present sequential data regarding adverse effects arising out of using the drug. Given the findings of our research, this is not surprising, since the use of sequential percentage data could facilitate more conservative updated judgments regarding likely adverse effects.

Current practices and our findings suggest that both marketers and regulators need to be careful about which data format to use when presenting a series or sequence of data. For instance, if consumers have self-positivity bias for an outcome (e.g., drinking and driving), then the use of frequency data for presenting sequential negative outcome data might be more appropriate to ensure that beliefs are normatively updated.

Finally, our findings suggest that encouraging careful processing of information (e.g., by increasing motivation or involvement) can actually reduce consumers' judgment accuracy in the case of sequential frequency data. This counterintuitive result suggests that well-intentioned strategies designed to get people to pay more attention to sequential data may in effect lead to more biased judgments. Further empirical work on this intriguing finding is warranted.

Limitations and Future Research Directions

Our studies were conducted in laboratory settings using college students. While we used realistic scenarios and products with which college students are familiar, extensions to other populations (possibly in field settings) are called for. In our studies, we deliberately limited the number of sequential data pieces to two to avoid participant fatigue. Future research might examine how participants update judgments with more than two pieces of sequential data. We examined scenarios of consistent ranking data based on a single dimension or attribute and with likelihoods greater than .50. What would be the effects when sequential rankings are mixed or based on multiple attributes (Dillon, Frederick, and Tangpanichdee 1985) or with likelihoods below .50? It might also be interesting to examine consumer cardinal ratings instead of the relative/ordinal judgments that we examine in this research. Another direction for research would be to examine processing of sequential data from online versus memory-based inputs (e.g., Jacoby et al. 2002). Also, future research should examine judgment updating when the data are in mixed formats—that is, when one datum is in frequency format and the other is in percentage format. Finally, there should be further research to extend our counterintuitive findings, whereby greater processing motivation led to poorer judgments. We hope this article encourages work in these and related areas.

REFERENCES

Anderson, N. H. (1981), *Foundations of Information Integration Theory*, New York: Academic Press.

Bar-Hillel, Maya (1980), "The Base-Rate Fallacy in Probability Judgments," *Acta Psychologica*, 44, 211–33.

Bearden, William, David M. Hardesty, and Randall L. Rose (2001), "Consumer Self-Confidence: Refinements in Conceptualization and Measurement," *Journal of Consumer Research*, 28 (June), 121–34.

Bennett, Peter D. and Gilbert D. Harrell (1975), "The Role of Confidence in Understanding and Predicting Buyers' Attitudes and Purchase Intentions," *Journal of Consumer Research*, 2 (September), 110–17.

Bialik, Carl (2006), "The Numbers Guy: Grading the Pollsters," *Wall Street Journal*, November 16, http://online.wsj.com/article/SB116360961928023945.html#mod=2_1125_1.

Birnbaum, Michael and Barbara A. Mellers (1983), "Bayesian Inference: Combining Base Rates with Opinions of Sources Who Vary in Credibility," *Journal of Personality and Social Psychology*, 45 (4), 792–804.

Brinberg, David, John G. Lynch, and Alan G. Sawyer (1992), "Hypothesized and Confounding Explanations in Theory Tests: A Bayesian Analysis," *Journal of Consumer Research*, 19 (September), 139–54.

Chatterjee, Subimal, Timothy Heath, Sandra Milberg, and Karen France (2000), "The Differential Processing of Price in Gains and Losses: The Effect of Frame and Need for Cognition," *Journal of Behavioral Decision Making*, 13 (1), 61–75.

Chen, Haipeng and Akshay Rao (2007), "When Two Plus Two Is Not Equal to Four: Errors in Processing Multiple Percentage Changes," *Journal of Consumer Research*, 34 (October), 327–40.

Cosmides, Leda and John Tooby (1996), "Are Humans Good Intuitive Statisticians after All? Rethinking Some Conclusions from the Literature on Judgment under Uncertainty," *Cognition*, 58, 1–73.

Dehaene, Stanislas (1992), "Varieties of Numerical Abilities," *Cognition*, 44, 1–42.

——— (1997), *The Number Sense: How the Mind Creates Mathematics*, New York: Oxford University Press.

——— (2001), "Précis of the Number Sense," *Mind and Language*, 16, 16–36.

DelVecchio, Devon, H. Shanker Krishnan, and Daniel C. Smith (2007), "Cents or Percent? The Effects of Promotion Framing on Price Expectations and Choice," *Journal of Marketing*, 71 (3), 158–70.

Diaconis, Persi and Sandy L. Zabell (1982), "Updating Subjective Probability," *Journal of the American Statistical Association*, 77 (380), 822–30.

Dillon, William R., Donald G. Frederick, and Vanchai Tangpanichdee (1985), "Decision Issues in Building Perceptual Product Spaces with Multi-attribute Rating Data," *Journal of Consumer Research*, 12 (June), 47–63.

Dougherty, Michael R. P. and James Shanteau (1999), "Averaging Expectancies and Perceptual Experiences in the Assessment of Quality," *Acta Psychologica*, 101 (1), 49–67.

Edwards, W. (1968), "Conservatism in Human Information Processing," in *Formal Representation of Human Judgment*, ed. B. Leinmuntz, New York: Wiley.

Erdem, Tülin, Michael P. Keane, T. Sabri Öncü, and Judi Strebel (2005), "Learning about Computers: An Analysis of Information Search and Technology Choice," *Quantitative Marketing and Economics*, 3 (3), 207–46.

Fox, Craig R. and Jonathan Levav (2000), "Familiarity Bias and Belief Reversal in Relative Likelihood Judgment," *Organizational Behavior and Human Decision Processes*, 82 (2), 268–92.

Garafalo, Joe and Frank K. Lester (1985), "Metacognition, Cognitive Monitoring, and Mathematical Performance," *Journal for Research in Mathematics Education*, 16 (3), 163–76.

Gershoff, Andrew D., Susan M. Broniarczyk, and Patricia M. West (2001), "Recommendation or Evaluation? Task Sensitivity in Information Source Selection," *Journal of Consumer Research*, 28 (December), 418–38.

Gershoff, Andrew D., Ashesh Mukherjee, and Anirban Mukhopadhyay (2003), "Consumer Acceptance of On-Line Agent Advice: Extremity and Positivity Effects," *Journal of Consumer Psychology*, 13 (1–2), 161–70.

Gigerenzer, Gerd and Ulrich Hoffrage (1995), "How to Improve Bayesian Reasoning without Instruction: Frequency Formats," *Psychological Review*, 102, 684–704.

——— (1999), "Overcoming Difficulties in Bayesian Reasoning: A Reply to Lewis and Keren (1999) and Mellers and McGraw (1999)," *Psychological Review*, 106, 425–30.

Gigerenzer, Gerd, Ulrich Hoffrage, and Heinz Kleinbölting (1991), "Probabilistic Mental Models: A Brunswikian Theory of Confidence," *Psychological Review*, 98 (4), 506–28.

Gollwitzer, Peter M. and Bernd Schaal (1998), "Metacognition in Action: The Importance of Implementation Intentions," *Personality and Social Psychology Review*, 2 (2), 124–36.

Gürhan-Canli, Zeynep (2003), "The Effect of Expected Variability of Product Quality and Attribute Uniqueness on Family Brand Evaluations," *Journal of Consumer Research*, 30 (June), 105–14.

Heath, Timothy, Subimal Chatterjee, and Karen R. France (1995), "Mental Accounting and Changes in Price: The Frame Dependence of Reference Dependence," *Journal of Consumer Research*, 22 (June), 90–97.

Hogarth, Robin M. and Hillel J. Einhorn (1992), "Order Effects in Belief Updating: The Belief-Adjustment Model," *Cognitive Psychology*, 24, 1–55.

Howard, John A. and Jagdish N. Sheth (1969), *The Theory of Buyer Behavior*, New York: Wiley.

Izard, Véronique and Stanislas Dehaene (2008), "Calibrating the Mental Number Line," *Cognition*, 106 (3), 1221–47.

Jacoby, Jacob, Maureen Morrin, James Jaccard, Zeynep Gürhan, Alfred Kuss, and Durairaj Maheswaran (2002), "Mapping Attitude Formation as a Function of Information Input: Online Processing Models of Attitude Formation," *Journal of Consumer Psychology*, 12 (1), 21–34.

Jadad, Alejandro R. and Anna Gagliardi (1998), "Rating Health Information on the Internet," *Journal of the American Medical Association*, 279, 611–14.

Johar, Gita V., Kamel Jedidi, and Jacob Jacoby (1997), "A Varying-Parameter Averaging Model of On-Line Brand Evaluations," *Journal of Consumer Research*, 24 (September), 232–47.

Juslin, Peter and Henrik Olsson (1997), "Thurstonian and Brunswikian Origins of Uncertainty in Judgment: A Sampling Model of Confidence in Sensory Discrimination," *Psychological Review*, 104 (2), 344–66.

Kahneman, Daniel and Amos Tversky (1972), "Subjective Probability: A Judgment of Representativeness," *Cognitive Psychology*, 3 (3), 430–54.

Kantowski, Mary Grace (1977), "Processes Involved in Mathematical Problem Solving," *Journal for Research in Mathematical Education*, 8 (3), 163–80.

Kichen, Steve and John J. Ray (2009), "Blue Chip Analysts: Analysts with the Best Results in the Worst of Times," *Forbes*, May 7, http://www.forbes.com/2009/05/07/zacks-stocks-analysts-brokerage-analysts-09_land.html.

Klayman, Joshua and Young-Won Ha (1987), "Confirmation, Disconfirmation, and Information in Hypothesis Testing," *Psychological Review*, 94 (2), 211–28.

Kopalle, Praveen and Donald R. Lehmann (2001), "Strategic Management of Expectations: The Role of Disconfirmation Sensitivity and Perfectionism," *Journal of Marketing Research*, 38 (August), 386–94.

Kruger, Justin and David Dunning (1999), "Unskilled and Unaware of It: How Difficulties in Recognizing One's Own Incompetence Lead to Inflated Self-Assessments," *Journal of Personality and Social Psychology*, 77 (6), 1121–34.

Lewis, Charles and Gideon Keren (1999), "On the Difficulties Underlying Bayesian Reasoning: A Comment on Gigerenzer and Hoffrage," *Psychological Review*, 106 (2), 411–16.

Lopes, L. Lola (1985), "Averaging Rules and Adjustment Process in Bayesian Inference," *Bulletin of the Psychonomic Society*, 23, 509–12.

McKenzie, Craig R. M. (1994), "The Accuracy of Intuitive Judgment Strategies: Covariation Assessment and Bayesian Inference," *Cognitive Psychology*, 26, 209–39.

Mellers, Barbara A., and A. Peter McGraw (1999), "How to Improve Bayesian Reasoning: Comment on Gigerenzer and Hoffrage," *Psychological Review*, 106, 417–24.

Metcalfe, Janet (2009), "Metacognitive Judgments and Control of Study," *Current Directions in Psychological Science*, 18 (3), 159–63.

Newell, Allen and Herbert A. Simon (1972), *Human Problem Solving*, Englewood Cliffs, NJ: Prentice-Hall.

Nickerson, Raymond S. (1998), "Confirmation Bias: A Ubiquitous Phenomenon in Many Guises," *Review of General Psychology*, 2 (2), 175–220.

Payne, John W. (1976), "Task Complexity and Contingent Processing in Decision Making: An Information Search and Protocol Analysis," *Organizational Behavior and Human Performance*, 16, 366–87.

Payne, John W., James R. Bettman, and Eric J. Johnson (1988), "Adaptive Strategy Selection in Decision Making," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14, 534–52.

——— (1993), *The Adaptive Decision Maker*, Cambridge: Cambridge University Press.

Pechmann, Cornelia and Chuan-Fong Shih (1999), "Smoking Scenes in Movies and Antismoking Advertisements before Movies: Effects on Youth," *Journal of Marketing*, 63 (July), 1–13.

Peters, Ellen, Daniel Västfjäll, Paul Slovic, C. K. Mertz, Ketti Mazzocco, and Stephan Dickert (2006), "Numeracy and Decision Making," *Psychological Science*, 17 (5), 407–13.

Peterson, Margaret Jean and Sonia Aller (1971), "Arithmetic Problem Solving," *Journal of Experimental Psychology*, 91 (1), 93–97.

Poletiek, Fenna H. and Mariëtte Berndsen (2000), "Hypothesis Testing as Risk Behavior with Regard to Beliefs," *Journal of Behavioral Decision Making*, 13 (1), 107–23.

Posavac, Steven S., David M. Sanbonmatsu, Frank Kardes, and Gavan J. Fitzsimons (2004), "The Brand Positivity Effect: When Evaluation Confers Preference," *Journal of Consumer Research*, 31 (December), 643–51.

Rust, Roland T., J. Jeffrey Inman, Jianmin Jia, and Anthony Zahorik (1999), "What You Don't Know about Customer Perceived Quality: The Role of Customer Expectation Distributions," *Marketing Science*, 18 (1), 77–92.

Sanbonmatsu, David M., Steven S. Posavac, Frank R. Kardes, and Susan P. Mantel (1998), "Selective Hypothesis Testing," *Psychonomic Bulletin and Review*, 5, 197–220.

Sanbonmatsu, David M., Steven S. Posavac, and Randon Stasney (1997), "The Subjective Beliefs Underlying Probability Overestimation," *Journal of Experimental Social Psychology*, 33, 276–95.

Schwarz, Norbert (2004), "Metacognitive Experiences in Consumer Judgment and Decision Making," *Journal of Consumer Psychology*, 14 (4), 332–48.

Shanteau, James (1975), "Averaging versus Multiplying Combination Rules of Inference Judgment," *Acta Psychologica*, 39 (February), 83–89.

Shiv, Baba, Julie A. Edell-Britton, and John Payne (2004), "Does Elaboration Increase or Decrease the Effectiveness of Negatively versus Positively Framed Messages?" *Journal of Consumer Research*, 31 (1), 199–208.

Shugan, Steve (1980), "The Cost of Thinking," *Journal of Consumer Research*, 7 (2), 99–111.

Simons, Katerina (1998), "Risk-Adjusted Performance of Mutual Funds," *New England Economic Review*, September/October, 1–16.

Slovic, Paul and Sarah Lichtenstein (1971), "Comparison of Bayesian and Regression Approaches to the Study of Information Processing in Judgment," *Organizational Behavior and Human Performance*, 6, 649–744.

Slovic, Paul, J. Monahan, and D. G. MacGregor (2000), "Violent Risk Assessment and Risk Communication: The Effects of Using Actual Cases, Providing Instructions, and Employing Probability versus Frequency Formats," *Law and Human Behavior*, 24 (3), 271–96.

Swaminathan, Vanitha (2003), "The Impact of Recommendation Agents on Consumer Evaluation and Choice: The Moderating Role of Category Risk, Product Complexity, and Consumer Knowledge," *Journal of Consumer Psychology*, 13 (1–2), 93–101.

Vinod, Hrishikesh D. (2004), "Ranking Mutual Funds Using Unconventional Utility Theory and Stochastic Dominance," *Journal of Empirical Finance*, 11 (3), 353–77.

West, Patricia M. and Susan M. Broniarczyk (1998), "Integrating Multiple Opinions: The Role of Aspiration Level on Consumer Response to Critic Consensus," *Journal of Consumer Research*, 25 (June), 38–51.