ANNE L. ROGGEVEEN and GITA VENKATARAMANI JOHAR*

The authors hypothesize that when managers integrate two projections to form a sales estimate, they evoke and use a sales range to judge inappropriately the plausibility of each projection. This judged plausibility, as well as the "margin of error" (based on the market research company's typical accuracy), is used to assign weights to each projection. Five experiments find strong evidence for this process and demonstrate a resulting bias.

Integration of Discrepant Sales Forecasts: The Influence of Plausibility Inferences Based on an Evoked Range

> When developing our business and marketing plans we did not anticipate spending hours pouring through conflicting research on market size and growth projections.
>
> —Steven Telio, Marketing Manager, MyHelpDesk.com, 2001

The epigraph illustrates an important problem in marketing: conflicting market research. This problem can have important implications for companies. For example, a marketing deal between SportsLine USA and America Online was stalled because of conflicting data on visitors to SportsLine's Web site (Green 1998). The problem of conflicting market research is widespread, often arising from differences in methodologies, definitions, and assumptions and affecting many forecasts, including market size and growth predictions, advertising reach numbers (Green 1998), and profitability predictions (Dempsey 1988). However, despite the discrepancies (or perhaps because of them), managers often solicit and receive multiple reports (Bunn and Salo 1993), recognizing that when multiple projections are combined appropriately, the resulting forecast can be more accurate than any of the individual ones (Falconer and Sivesind 1977).

*Anne L. Roggeveen is Assistant Professor of Marketing, Babson College (e-mail: aroggeveen@babson.edu). Gita V. Johar is Professor of Marketing, Columbia Business School, Columbia University (e-mail: gvj1@columbia.edu). This article is based on the first author's doctoral dissertation conducted under the guidance of the second author. The authors thank the dissertation committee members, Andrew Gershoff, Don Lehmann, Vicki Morwitz, and Elke Weber, for their constructive suggestions at all stages of this research. The authors also thank Tim Heath, Erik Rijnbout, Jaideep Sengupta, Chuck Weinberg, Klaus Wertenbroch, and the three anonymous *JMR* reviewers for their helpful comments and suggestions. This research was supported by the Columbia Business School Research Fund and the Babson Board of Research Fund.

Forecasting is an extremely important part of marketing planning. Of 134 companies surveyed, 99% indicated that they prepare forecasts when creating their marketing plans (Dalrymple 1987), and many high-level executives believe that forecasting is synonymous with planning (Lapide 2000). Because of the importance of the forecast in marketing planning, an understanding of how managers estimate sales when faced with discrepant forecasts from different market research sources is critical.

Much research has been devoted to describing forecast methods and highlighting conditions in which each method is most accurate and useful (Armstrong 2001). Forecasting methods can be categorized into two broad categories: (1) those based primarily on judgment (e.g., Mentzer and Kahn 1995; Sanders and Manrodt 1994) and (2) those based on statistical sources. Research on forecast combination has focused on the delineation of normatively appropriate ways to combine forecasts and has acknowledged that more work is needed to describe the actual integration process that decision makers use (Clemen 1989; Flores and White 1988). The goal of this article is to describe this process and highlight the potential deviation of the outcome from a normative benchmark.

We describe five experiments that show that when a second projection (1) is lower than a first projection and (2) has a higher margin of error than the first projection, it receives more weight than it normatively should. However, the second projection receives the "normatively correct" amount of weight when it is higher than the first. Weight given to each forecast should be determined only by the market research company's typical forecast accuracy, as described by its margin of error; however, managers use the range in which the forecasts are expected to lie as additional input to infer the accuracy of each forecast. Use of this information distorts the process. Reliance on the range and the way it affects the integration process is explained by range-frequency theory.

LITERATURE REVIEW

Combining Forecasts: The Norm

It is widely acknowledged that forecast combination constructed from several separate sources leads to increased forecast accuracy (e.g., Batchelor and Dua 1995; Clemen 1989). A large body of research has therefore examined the "right" way to combine independent forecasts (for extensive reviews, see Armstrong 2001; Clemen 1989). Various approaches are available for combining forecasts, including minimum-variance models (e.g., Kang 1986), regression models (e.g., Granger and Ramanathan 1984), Bayesian models (e.g., Morris 1977), econometric models (e.g., Schmittlein, Kim, and Morrison 1990), and combining probability distributions (for a review, see Genest and Zidek 1986). A simpler approach that has proved relatively robust and more accurate than other more elaborate schemes in many applications is to take an equally weighted arithmetic average of the individual forecasts (Figlewski and Urich 1983; Makridakis and Winkler 1983). However, an even better-performing model than the simple averaging model (but still not too elaborate) is one in which the weighting scheme for divergent forecasts is based on subjective assessments of the precision of the forecasters (Ashton and Ashton 1985). A relatively easy and normatively appropriate procedure of combining multiple forecasts is to weight each forecast relative to the market research company's typical accuracy, as indicated by its margin of error. However, the decision-making literature on context effects suggests that decision makers' forecast weights are likely to be influenced by contextual factors as well.

Combining Forecasts: Role of Context

Judgments of forecasts are not made in a vacuum; for existing products, there is knowledge of past sales, and for both existing and new products, there is knowledge of the marketplace. This background knowledge is likely to suggest a distribution of possible sales, which will then provide a frame of reference each time a judgment regarding forecasts is made (Kahneman and Miller 1986). We hypothesize that the evoked range makes some forecasts appear more plausible than others, which in turn causes people to make inferences about the forecasts' accuracy. This then affects the weight that is applied to each forecast in the integration process. In actuality, the range in which forecasts can lie (e.g., from orders received thus far to total market size) does not provide information about the accuracy of the individual forecasts; only the margin of error around each forecast provides that. However, inferences about which forecast is more accurate and thus should be weighted more may be drawn simply on the basis of where the forecast lies in the evoked range. This article focuses on this nonnormative integration process.

How can evoked ranges influence forecast integration? According to range-frequency theory (Lim 1995; Niedrich, Sharma, and Wedell 2001; Parducci 1965, 1995), which emphasizes the contextual nature of perception and judgment, the judged value of a stimulus (e.g., a forecast) is a joint function of two aspects of the evoked distribution. The first aspect, the "range principle," specifies that judgments about the value reflect the location of the target stimulus relative to the highest and lowest values in the distribution (Janiszewski and Lichtenstein 1999; Nunnally 1978; Sherif and Hovland 1961; Volkmann 1951; Wedell and Parducci 1988). For example, Janiszewski and Lichtenstein (1999) demonstrate that the range of prices a consumer evokes when evaluating a market price can independently influence judgments about the attractiveness of the market price. They find that consumers judge a price of $1.25 as attractive when it is nearer to the lower boundary of known prices ($.99–$1.74), as moderately attractive when it is the midpoint of known prices ($.74–$1.74), and as unattractive when it is nearer to the upper boundary of a set of known prices ($.74–$1.49).

The second aspect, the "frequency principle," specifies that judgments of a target stimulus depend on the relative frequencies of other stimuli that are present in the distribution. For example, under the assumption of a single range from $5 to $15, consumers would consider $10 more expensive after buying items for $6 and $7 than they would after buying items for $13 and $14. In other words, when most of the stimuli lie below a target stimulus, the target appears to have a higher standing than when most stimuli lie above the target. The subjective judgment of a stimulus is conceived as a compromise between the range and frequency principles, which reflects the relative importance of each principle in a given situation.

Under the assumption that the range-frequency principle applies to the integration of discrepant forecasts, the central question is, What range is likely to be evoked during the integration process? Consider the case of really new products, for which the manager has no information about prior sales of similar products. In this case, the easiest low number to generate is "zero sales." This is consistent with literature that suggests that the mental number line begins at zero or one (Dehaene 1997; Devlin 2000); however, there is no point at which the mental number line ends (Dehaene 1997; Devlin 2000). Thus, there is no firmly established internal anchor to serve as an upper endpoint for a plausible sales range. In the absence of any information, the upper bound is likely to be the highest sales projection received. This proposition is based on research that demonstrates that internal anchors assimilate to proximate external stimuli, especially when the internal anchors are not firmly established (see Lichtenstein and Bearden 1989; Sherif and Hovland 1961; Urbany, Bearden, and Weilbaker 1988).

In the case of existing products, the previous year's sales and market size can be lower and upper bounds of the range of possible sales. If just one projection (Projection A) is received, the evoked range is likely to be from either the previous year's sales (presumed to be lower than Projection A) or zero to Projection A (if the market size is unknown). Where the second projection (Projection B) lies in relation to the distribution influences how the two projections are evaluated. If a stimulus (Projection B) is added beyond the upper end (Projection A) of the distribution (as is the case when Projection B is higher than Projection A), the lower end of the distribution (zero or previous sales) seems significantly lower than the two projections, thereby causing a contrast effect (Cooke et al. 2002). This results in the lower end of the distribution seeming less plausible and the upper end (from Projection A to Projection B) seeming more plausible. In this case, the relevant range is considered Projection A to Projection B. When the range is anchored by the two forecasts, each forecast seems equally plausible on the basis of the range; thus, the range does not provide information about the relative weighting scheme. The only diagnostic information about the relative weighting of each forecast arises from the respective margins of error of each forecast. Use of only this information is what the normative approach dictates and is therefore likely to result in the normative estimate.

If a stimulus (Projection B) is inserted into the existing distribution (as is the case when Projection B is lower than Projection A), neither endpoint will seem particularly extreme. Thus, neither endpoint will be contrasted out. However, in terms of the two projections, Projection A, which represents the upper bound of the evoked range, will be viewed as an extreme case. Projection B, which lies within the range, will be viewed as a more likely scenario. In this case (i.e., Projection B < Projection A), when the margin of error for Projection B is greater than that of Projection A, the plausibility cue derived from range suggests a different weighting than that suggested by margins of error. We expect that this results in Projection B receiving more weight in the integration than it should, solely according to its margin of error, which possibly results in a nonnormative estimate. Note that the preceding discussion implies that Projection B will be weighted more heavily when it is lower than Projection A than when it is higher than Projection A.

We tested these notions by setting up situations in which the margin of error of the first projection (A) is always less than the margin of error of the second projection (B), so that normatively Projection A should be weighted more heavily than Projection B.[1] The plausibility cue inferred from the evoked range is either nondiagnostic (when B > A) or inconsistent (when B < A) with the weighting scheme suggested by the margins of error. When the plausibility cue is nondiagnostic, the weighting scheme should respect the relative margin of errors and result in an estimate that is close to normative. When the plausibility cue and the margins of error are inconsistent (when B < A), Projection B being weighted more heavily than Projection A (when it actually has a higher margin of error than Projection A) will provide a strong test of the theory.

Two alternative processes (the negativity bias and prior beliefs) also predict that a second lower sales forecast is weighted more heavily than a higher one during integration. In general, people weight negative information more heavily than positive information (Fiske 1980; Skowronski and Carlston 1989). In the forecasting context, such a process might reflect conservatism. A second motivational process that would also predict the overweighting of lower forecasts is a belief that market research is inflated (e.g., Cooper and Consoli 1999). We test these two motivational explanations against the evoked-range explanation in Experiment 4.

[1]We ran one experiment in which margin of error for Projection A (18%) was higher than margin of error for Projection B (10%). When Projection B is higher than Projection A, integration is based only on respective margins of error, as we expected. When Projection B is lower than Projection A, the plausibility cue and the margin of error cue both point to downward adjustment from Projection A, and this consistency in the cues should also result in a greater movement toward Projection B when it is lower than Projection A than when it is higher than Projection A. Our results support this proposition.

PLAN OF STUDIES

Similar to Schmittlein, Kim, and Morrison (1990), we focus on combining two forecasts. Experiment 1 demonstrates the outcome of the posited range-frequency process when no other information is available. As we predict, there is more integration toward a second lower sales forecast than toward a second higher forecast; furthermore, the weighting of the second lower forecast is nonnormative and contrary to the respective margins of error. This occurs in both low- and high-accountability conditions. Experiment 2 replicates the result in a different decision-making domain, even when participants believe that the costs of underforecasting are high. Experiment 3 replicates the result by means of a market size estimation task in a simulated business environment in which previous market sizes are known. Experiment 4 rules out the negativity bias and prior beliefs explanations. Experiment 5 provides direct evidence for the role of evoked ranges by demonstrating that the plausibility of a projection within a range determines how much it is weighted.

The general stimuli and method are similar across Experiments 1, 4, and 5. The experimental stimuli consist of two sales-unit projections (units delivered) for a fictitious Internet start-up company that offers an online grocery-delivery service (see Table 1). We refer to the projections as Projection A (the initial projection, constant across participants) and Projection B (the second projection, manipulated to be lower or higher than Projection A). The data collection procedures used by the two different companies that provided the projections were similar, and the company names were fictitious and rated equally credible in a pretest. Across all experiments (except Experiment 1D), we determined Projection B by subtracting 70% of Projection A from Projection A in the lower condition and by adding 70% of Projection A to Projection A in the higher condition. Thus, the difference between Projections A and B was identical, regardless of whether Projection B was lower or higher than Projection A. Participants were randomly assigned to the lower or higher Projection B condition. We completed all three experiments before the failure of online grocery services such as Webvan. To reduce the likelihood that participants would have prior beliefs about the area in which the company offered the service, we selected a start-up company located in a state about which we expected that participants had no prior knowledge.

Participants were asked to imagine that they were managers at a start-up online grocery service. They were told that for business-planning purposes, they needed to forecast unit sales. To aid in this, two market research companies had been hired to provide projections. Since the time the work was commissioned, it had been discovered that the market research company that provided Projection A typically had a margin of error of ±2%. All participants were told that the market research company that provided Projection B typically had a margin of error of ±10%. Margin of error was explained and typical error ranges were given as between ±2% and ±18%. The two projections of annual unit sales were given to participants along with the margin-of-error rates for each projection. Participants were then asked to provide their sales estimates. They were allowed to look at the research reports when making their estimates.

Table 1

MANIPULATED FACTORS AND PROJECTIONS (MARGIN-OF-ERROR RATES) GIVEN IN EXPERIMENTS 1–2 AND 4–5

Experiment	*Manipulated Factors (Constant Factors)*	*Projection A Constant Across Conditions*	*Projection B Condition: Lower Than A*	*Projection B Condition: Higher Than A*
1	B is lower/higher than A Accountability	651,600 (±2%)	195,480 (±10%)	1,107,720 (±10%)
1A	B is lower/higher than A (Range given with endpoints slightly below/above A and B)	3,733,000 (±4%)	1,119,900 (±10%)	6,346,100 (±10%)
1B	B is lower/higher than A (Give estimates after A and after B)	528,900 (±2%)	158,670 (±10%)	899,130 (±10%)
1C	B is lower/higher than A (Projections on same page)	528,900 (±2%)	158,670 (±10%)	899,130 (±10%)
1D	B is lower/higher than A (Ratio between A and B constant)	528,900 (±2%)	158,670 (±10%)	1,763,000 (±10%)
2	B is lower/higher than A Costs of under- and overestimation	528,900 (±2%)	158,670 (±10%)	899,130 (±10%)
4	B is lower/higher than A Context (research/audit) Frame (low number good/bad)	528,900 (±2%)	158,670 (±10%)	899,130 (±10%)
5	B is lower/higher than A Explicit range given	3,733,000 (±4%)	1,119,900 (±10%)	6,346,100 (±10%)
5A	B is lower/higher than A (Verbal protocols)	528,900 (±2%)	158,670 (±10%)	899,130 (±10%)

Notes: All participants saw two projections: A and B. Projection A was the same for all participants, and we manipulated Projection B such that half the participants received a second projection that was lower than Projection A and half the participants received a second projection that was higher than Projection A.

EXPERIMENT 1: DEMONSTRATION OF BIASED INTEGRATION

We designed Experiment 1 to examine the effect of type of information (i.e., whether Projection B, the manipulated projection, is lower or higher than Projection A) on integration in a situation in which no other contextual information (e.g., prior sales) is present. As we described previously, we expect that Projection B is weighted more heavily during integration when it is lower (rather than higher) than Projection A. We also manipulated accountability to underscore the perceptual (rather than motivational) nature of the integration process. The existing body of research on the effects of making people accountable for their judgments (Janis and Mann 1977; Tetlock and Boettger 1989) suggests that people who expect to justify their views later to an unknown audience are highly concerned with potential errors in their judgments because of fear of embarrassment and loss of self-esteem in the event of delivering a "bad" judgment (Janis and Mann 1977; Tetlock 1983). Therefore, accountable subjects should be motivated to make accurate judgments and to be less prone to errors in integration. However, range-frequency theory relies on a perceptual process, which suggests that accountability should not entirely remove the effect.

Method

Participants and design. Participants were 58 students who were paid $8 each for their participation in the 2 × 2 between-subjects experiment. The between-subjects factors were accountability (low versus high) and whether Projection B was lower or higher than Projection A.

Procedure. Participants received two packets, each of which contained a market research report with the unit-sales projection and its associated margin-of-error rate. Depending on the accountability manipulation, participants were told one of two things (Tetlock and Boettger 1989). In the low-accountability condition, participants were told the following: "There is no way for us to associate your individual responses with you. Therefore, your answers will be kept completely confidential and anonymous. No one will know how you respond." In the high-accountability condition, participants were told the following: "After you finish this exercise, you will go to another room where you will be interviewed about the exercise. In the interview you will be asked to explain the information you used and why you answered the question the way you did. Please sign below, acknowledging that we may audiotape the follow-up interview for future data-analytic purposes." Participants then signed and dated the sheet. Participants read the market research report with Projection A and then opened the second packet, which reminded them of the accountability manipulation and contained the market research report with Projection B (lower or higher than Projection A). They then made their estimates.

Results

Sales estimates. To determine the impact on the sales estimate of Projection B being lower or higher than Projection A, we used the absolute value of the difference between the participant's estimate and Projection A as the dependent measure. An analysis of variance (ANOVA) revealed a significant main effect of whether Projection B was lower or higher than Projection A (M lower = 352,548, M higher = 115,911; $F(1, 51) = 17.17$, $p < .001$). When Projection B was lower, participants adjusted more toward that projection than when Projection B was higher. No other effects were significant (all ps > .7).

Comparison to norm. Do people overadjust toward the lower projection or underadjust toward the higher projection? A method for determining a normative response is to use a weighted average of the individual projections (Stone 1961). The weights applied to each forecast should be based on the confidence in the given information (Armstrong 2001). When margin of error is known, this confidence is determined by (1/margin of error), and the normative weight for new (initial) information is determined as follows:

$$(1) \quad \frac{\text{(1/margin of error of new [initial] information)}}{\text{[(1/margin of error of initial information)} + \text{(1/margin of error of new information)]}}.$$

Given the provided margin-of-error rates (2% and 10%), the weight that participants should have applied to the new information was .167, resulting in a normative response of 575,428 when Projection B was lower and 727,772 when Projection B was higher. In both the low- and the high-accountability conditions, the estimate participants provided was significantly lower than the normative response when Projection B was lower (low accountability = 293,756, $t[13] = -8.46$, $p < .005$; high accountability = 304,348, $t[13] = -6.27$, $p < .005$) and at the normative response when Projection B was higher (low accountability = 754,831, $t[12] = .36$; high accountability = 779,286, $t[13] = .75$).

Discussion

Consistent with the hypothesis about an evoked-range process, participants adjusted more toward Projection B when it was lower (rather than higher) than Projection A. This result obtained even when participants believed that they were accountable, which suggests that it is not a purely motivational tendency. The amount of adjustment was non-normative in the "B-lower" condition and normative in the "B-higher" condition. Under the assumption that there is an averaging integration rule, the weight given to Projection B when it is lower than Projection A (accountability: low w_B = .78, high w_B =.76) not only is significantly different from the normative weight (.167) but also is greater than the weight given to Projection A (accountability: low w_A = .22, high w_A = .24). This occurs even though Projection A's margin of error (2%) is lower than Projection B's margin of error (10%). This strong result suggests that the effect of evoked range on perceived accuracy of each forecast overwhelms the influence of the reliability of each forecast in this case.

We ran a series of follow-up experiments to test the hypothesized evoked-range process. Experiment 1A determined that when subjects are given an explicit range in which the lower endpoint is slightly lower than the lowest projection seen and the upper endpoint is slightly higher than the highest projection seen, movement from Projection A is the same whether Projection B is lower (M = 835,028) or higher (M = 867,000) than Projection A. Furthermore, estimates participants provided are normative even when Projection B is lower than Projection A (lower normative = 2,985,653, lower estimate = 2,897,972, $t[19] = .48$; higher normative = 4,480,347, higher estimate = 4,600,000, $t[17] = 1.03$). In this case, the range does not provide information about the relative plausibilities of the two forecasts; thus, only one cue is used: respective margins of error. The evoking of a range and use of it to judge forecast plausibility appears to be responsible for the overweighting of a second lower forecast in Experiment 1.

To determine whether participants evoke information other than the forecast provided when no other information is present, we ran Experiment 1B, which asked participants to estimate sales immediately after they viewed Projection A; estimates were significantly below Projection A (given Projection A = 528,900; estimate = 466,522, $t[26] = -3.02$, $p < .005$), indicating that participants did not simply rely on the provided projection but used additional input in arriving at their estimate. Estimates provided after exposure to Projections A and B replicate results from Experiment 1 (movement from Projection A: M lower = 268,115, M higher = 61,933). Experiment 1C used projections listed on the same page, with Projection A listed above B; again, the results replicated (M lower = 130,840, M higher = 43,227). Finally, Experiment 1D held the ratio between Projections A and B constant (instead of holding the difference constant, as in other experiments); there was greater proportional adjustment toward Projection B when it was lower than Projection A than when it was higher than Projection A (M lower = .46, M higher = .12). Again, the adjustment toward a lower (but not higher) Projection B was nonnormative.

Experiments 1A–D demonstrate a robust tendency of integration that is nonnormative when a second projection received is lower than the first. However, the stimuli were rather pallid, and participants did not have any information other than the forecasts. It could also be argued that the tendency to overweight lower forecasts does not occur when the costs of forecasting are salient. Experiment 2 tests the hypothesis with these issues in mind.

EXPERIMENT 2: INCORPORATION OF COSTS OF OVER- AND UNDERFORECASTING

Experiment 1 did not provide participants with information about the costs associated with inaccurate forecasting. In the real world, when managers make forecasts, there are costs associated with over- or underforecasting sales. For example, overforecasting can result in excess inventory, and underforecasting can result in the loss of revenue to competition. Explicit consideration of these costs is likely to affect a manager's forecast. If an "accuracy motivation" mechanism underlies the tendency to overweight a second forecast when it is lower than the first, this bias may not occur when the cost associated with underforecasting is made salient. The proposed perceptual mechanism suggests that the bias will manifest regardless of costs.

Method

Participants and design. Participants were 59 students who were paid $6 to participate in a 2 × 2 between-subjects experiment. The between-subjects factors were Projection B being lower or higher than Projection A and costs (underestimating is high versus overestimating is high).

Procedure. Participants were asked to play the role of a manager at a Honda dealership. They were provided market research information about Civic Hybrid cars and were asked their opinion on the number of Civic Hybrids that would be sold between 2003 and 2008. They were told that this information would be used in determining the level at which to set plant capacity and to read the information about the cost of under- or overforecasting (depending on their condition).

All subjects then read a presentation about the Honda Civic Hybrid, which included a definition and description of how the car worked, its fuel efficiency, and other car facts. They also read information about how the media describe the car and its competition. Then, participants received two sales projections and their associated margin-of-error rates (see Table 1) and provided their sales estimates.

Manipulations. In the high-cost-of-underestimating condition, participants were told the following: "If you underforecast sales (i.e., forecast below what will actually occur), the plant will not be able to handle the capacity. Customers will not wait for back orders to be produced; they will simply purchase from the competition. This will reduce Honda's profitability. Therefore, when providing your opinion, you should keep in mind that the potential costs of underforecasting sales (forecasting below what will actually occur) are high." In the high-cost-of-overestimating condition, participants were told the following: "If you overforecast sales (i.e., forecast above what will actually occur), the plant will produce excess inventory. The company will have to pay the cost of storing the excess cars and will have to sell the excess cars at discount to reduce inventory. This will reduce Honda's profitability. Therefore, when providing your opinion, you should keep in mind that the potential costs of overforecasting sales (forecasting above what will actually occur) are high."

We tested the manipulations in a pilot study of 57 subjects who saw one of the two cost manipulations and then responded to two questions (on seven-point scales) about the cost of forecasting. The results show that the manipulations worked as intended. There was greater agreement that the cost of overforecasting was high in the overforecast condition than in the underforecast condition (M over = 5.5, M under = 4.9; $F(1, 53) = 3.91$, $p = .05$). Similarly, in the underforecast (versus overforecast) condition, there was greater agreement that the cost of underforecasting was high (M over = 3.8, M under = 5.4; $F(1, 53) = 22.13$, $p < .001$). Note that participants perceived the magnitude of costs similarly in the two conditions (M over = 5.5, M under = 5.4; $F < 1$).

Results

Sales estimates. An ANOVA on the amount of movement revealed a significant main effect of type of Projection B ($F(1, 55) = 17.86$, $p < .001$). There was greater integration of Projection B when it was lower than Projection A (M = 217,633) than when it was higher than Projection A (M = 58,495). The interaction between type and cost was not significant ($F < 1$).

Comparison to norm. The estimate participants provided was significantly lower than the norm when Projection B was lower (normative = 467,072; high cost of underestimating = 346,350, $t[13] = –3.04$, $p < .005$; high cost of overestimating = 276,183, $t[14] = –6.15$, $p < .005$) and at the norm when Projection B was higher in both cost conditions (normative = 590,728; high cost of underestimating = 559,603, $t[14] = –.68$; high cost of overestimating = 615,187, $t[14] = .74$). Again, weight given to the second forecast (Projection B) when it was lower was either the same as or more than the weight given to the first forecast (Projection A) (high cost of underestimating $w_B = .49$, $w_A = .51$; high cost of overestimating $w_B = .68$, $w_A = .32$), even though the second forecast had a greater margin of error.

Discussion

Results demonstrate that the tendency to overweight lower sales forecasts occurs even when the information provided is detailed and set in a decision-making context. Furthermore, this tendency to underforecast sales significantly occurs even when participants are explicitly told that the cost of underforecasting sales is high. The effect remains the same regardless of whether the cost of under- or overforecasting is salient. This suggests a robust perceptual phenomenon that manifests regardless of motivation. Although the demonstration of the effect thus far is compelling, it could be argued that we obtained the results from students, who had limited experience in dealing with business forecasts; the results may not replicate with true managerial judgments. We ran Experiment 3 to address this issue.

EXPERIMENT 3: INTEGRATION IN MANAGERIAL SETTINGS

We ran Experiment 3 to determine whether managers exhibit the same pattern of integration in a setting in which they are familiar with the data and the business. Another purpose of Experiment 3 was to replicate the results in a scenario in which contextual information could provide input to the evoked range (i.e., the lower bound need not be assumed to be close to zero but is explicit).

We conducted Experiment 3 in the context of MARKSTRAT3 (Larreche and Gatignon 1998), a business simulation game that requires users to analyze information and make strategic decisions in a series of time periods. In general, MARKSTRAT is considered among the most realistic business simulations (Lambert 1980) and has provided data for research in the area of forecasting and decision making (Glazer, Steckel, and Winer 1989; Hogarth and Makridakis 1981). Replication of the results in this scenario would generalize the result to situations in which experience with the market has accumulated over time.

Method

Participants and design. Participants were 37 MBA and executive MBA students who had at least four years of business experience and at least two semesters of MBA work and had been working with the MARKSTRAT3 simulation over a seven-week period. To encourage participation, we entered completed surveys into a lottery to win a bottle of wine. The experiment used a between-subjects design, and type of Projection B (lower versus higher than A) was the between-subjects factor.

Procedure. After participants turned in their decisions for Period 6, they were given a survey that explained that there were two market research companies available in the MARKSTRAT3 simulation, but that for simplicity they had only been working with projections provided by Research Company X. The survey then gave projections and the associated margins of error from both Research Company X (4% margin of error) and Research Company Y (10% margin of error) and asked participants to estimate individually the actual market sizes for two products.

Stimuli. The experimental stimuli consisted of market research results, from two companies, about the forecasted size of the MARKSTRAT3 markets (Sonite and Vodite). Participants were provided three numbers: the actual market size in Period 6 (from the simulation; e.g., Industry 1, Sonite = 1,750,000) and two projections. Projection A (e.g., 1,786,000), attributed to Research Company X, was the projection that the MARKSTRAT3 simulation gave for Period 7 and was always higher than the actual market size in Period 6. We manipulated Projection B, attributed to Research Company Y. We determined Projection B by either subtracting or adding 70% of the difference between the actual market size in Period 6 and Projection A to Projection A (e.g., lower = 1,760,800; higher = 1,811,200). Thus, Projection B was always higher than the actual market size in Period 6.

The simulation was played in five industries with three to five teams per industry. Each team had two members on average. Each person was asked to fill out a survey independently. One person on the team knew the actual market size for Period 6, Projection A (from the simulation), and the lower Projection B. Another person on the team knew the same three pieces of information, except he or she knew the higher Projection B. Each industry in the simulation had a different market size in Period 6 and a different projected market size in Period 7; therefore, the numbers differed across industries.

Results

Overview. Because the given information for the Sonite and Vodite markets varied across industries, for each participant, we calculated a ratio of absolute value of the difference between the participant's forecast and Projection A to total possible difference (absolute value of the difference between Projections B and A). This variable served as the dependent measure.

Market size. An ANOVA on the ratio of movement in market size for the two products (Vodites and Sonites) as a within-subjects factor with type of Projection B (lower versus higher than Projection A) as a between-subjects factor revealed a significant main effect for type of Projection B. As we expected, there was more movement toward the lower (rather than higher) projection (Sonite: M lower = .72, M higher = .23, Vodite: M lower = .55, M higher = .31; $F(1, 35) = 11.26$, $p < .001$). Product had no main or interactive effects, which suggests that the tendency to move more when the second projection is lower (rather than higher) than the first generalizes across the two products.

Comparison to norm. For both Sonites and Vodites, the estimate participants provided was significantly lower than the normative response when Projection B was lower (normative = .236; Sonite = .72, $t[18] = 3.32$, $p < .005$; Vodite = .55, $t[18] = 2.62$, $p < .01$) and at the normative response when Projection B was higher (normative = .236; Sonite = .23, $t[17] = -.77$; Vodite = .31, $t[17] = .36$).

Discussion

Experiment 3 replicates previous results and demonstrates that MBAs, working with projections in a simulation that they had used over an extended period, adjust more toward Projection B when it is lower (rather than higher) than Projection A; furthermore, estimates are nonnormative when Projection B is lower than Projection A. It could be argued that using a weighted average based on the reliability of each report might not be an appropriate normative standard in this situation because participants knew how the forecasts related to actual market size in previous periods and correctly used that information as additional input. Historically, the Sonite market-size forecasts varied between higher and lower than actual sales, and thus previous knowledge did not provide diagnostic information. In two of the industries in which the forecasts historically had been too pessimistic (and thus subjects should have given more credence to the new forecast when it was higher), there was still a pattern of adjusting more toward a new, lower forecast. In the Vodite market, the market-size forecasts historically provided in the simulation had been optimistic. However, even in industries in which recent Vodite forecasts had been too low (and thus subjects should have given more credence to the new forecast when it was higher), the pattern of adjusting more toward the new, lower forecast replicated.

In summary, even experienced decision makers were subject to a perceptual distortion in the integration process, brought about by considering the forecasts in the context of a possible range. An alternative explanation for the robust pattern of results is that the new, lower projection is weighted more heavily because it implies bad news for the company, and managers want to be conservative in forming a forecast. In this experiment, however, because the dependent variable was the total market size and not individual company sales, any motivational pressure to underestimate should have been reduced. Experiment 4 examines the conservatism explanation more closely.

EXPERIMENT 4: IS IT A CONSERVATISM BIAS?

Results from Experiments 1–3 do not imply a conservatism explanation, because if conservatism held, integration should have been biased in the direction of the lower forecast, regardless of whether it was the first or second forecast that participants received. However, we consistently found that integration is nonnormative only when the second forecast is lower than the first, but it is possible that exposure to a second lower forecast stimulates conservatism. To test this explanation, Experiment 4 orthogonally manipulates the meaning of the second forecast (Projection B) (good news versus bad news) with its value (lower/higher than Projection A).

Experiment 4 also examines whether a belief that market research is inflated is stimulated when a second lower forecast is received, thereby resulting in the overweighting of that forecast. We examined this by manipulating context, that is, whether the sales projections are stated to be from a market research agency or an audit company. We selected an audit context because, in general, audit results at the time we conducted the experiment (before the recent corporate accounting/auditing scandals) were considered precise, reliable, and trustworthy (Singleton-Green 1990).

Method

Participants and design. Participants were 120 undergraduate and graduate students who were paid $12 each for their participation. The study was a 2 × 2 × 2 between-subjects design. Projection B compared with Projection A (lower versus higher), context (market research versus audit), and frame (lower number implies bad news versus lower number implies good news) were the factors.

Stimuli and procedure. The experimental procedure was similar to that of Experiment 1 in the market research context (for stimuli, see Table 1). In the audit context, participants were asked to imagine that they were managers at a start-up company that offered an online grocery service and was interested in acquiring a smaller online grocery business. Two audit firms provided projections of the smaller online grocery service. In all context conditions, we framed the projections provided in one of two ways. In one frame, a lower number implied bad news (i.e., units delivered), and in the other frame, a lower number implied good news (i.e., delivery failures). For both frames, we provided the same projections. If the weighting of new, lower information is driven by the implication of bad news for the company (i.e., conservatism), it should be weighted more heavily for units delivered; the new, higher projection should be weighted more heavily for delivery failures. The proposed perceptual evoked-range process suggests that results will replicate in the low-number-is-good-news and low-number-is-bad-news conditions.

Results

Manipulation checks. Participants were asked whether Projection B represented bad news (1) or good news (7). An ANOVA revealed a main effect of frame (M deliveries = 4.32, M delivery failures = 3.66; $F(1, 110) = 7.33$, $p < .01$). Notably, there was also a significant interaction between type of Projection B and frame ($F(1, 110) = 38.87$, $p < .001$). Follow-up contrasts revealed the expected pattern such that when Projection B indicated more deliveries, participants viewed that information as better news than when Projection B indicated fewer deliveries (M lower = 3.57, M higher = 5.10; $F(1, 114) = 20.06$, $p < .001$). When Projection B indicated more delivery failures, participants viewed that information as worse news than when Projection B indicated fewer delivery failures (M lower = 4.40, M higher = 2.90; $F(1, 114) = 19.20$, $p < .001$).

Sales estimates. The ANOVA on the amount of movement revealed a significant main effect of type of Projection B (M lower = 223,122, M higher = 115,283; $F(1, 107) = 14.74$, $p < .001$). Participants integrated Projection B more when the second projection was lower (rather than higher); this occurred regardless of whether a lower forecast implied bad news (fewer deliveries) or good news (fewer delivery failures; interaction $p > .4$). Other effects, including ones involving context (market research versus audit), were not significant.

Comparison to norm. In all conditions, the estimate participants provided was significantly lower than the norm when Projection B was lower than Projection A, regardless of whether the lower projection implied good news or bad news for the company (normative = 467,072; M lower is bad, research = 254,983, $t[14] = –5.07$, $p < .005$; M lower is good, research = 334,976, $t[13] = –3.85$, $p < .005$; M lower is bad, audit = 345,853, $t[14] = –3.93$, $p < .005$; M lower is good, audit = 289,245, $t[14] = –5.88$, $p < .005$). The estimate was at the norm when Projection B was higher than Projection A for all conditions except the audit/delivery failures scenario in which the response was above the norm. In this condition, participants adjusted more toward the negative information (higher delivery failures) than they normatively should have, which suggests some role for a negativity bias (normative = 590,728; M lower is bad, research = 586,500, $t[13] = –.08$; M lower is good, research = 652,241, $t[12] = 1.63$; M lower is bad, audit = 661,942, $t[14] = 1.51$; M lower is good, audit = 675,357, $t[13] = 2.52$, $p < .025$).

Discussion

Once again, in a context in which no information other than the projections themselves is available, new, lower numbers carry greater weight than do new, higher numbers. The normative weight of .17 for the second forecast (Projection B; which is significantly lower than the normative weight for Projection A because of the associated margins of error) is applied when Projection B is higher than the initial forecast (Projection A). However, when Projection B is lower than Projection A, either the same weight or more weight is given to the second forecast than to the first forecast (lower is bad, research $w_B = .74$, $w_A = .26$; lower is good, research $w_B = .52$, $w_A = .48$; lower is bad, audit $w_B = .49$, $w_A = .51$; lower is good, audit $w_B = .65$, $w_A = .35$). In all cases, the resultant integrated estimate is significantly different from the normative estimate. This occurs regardless of whether the lower forecast implies good news or bad news for the company. Furthermore, results indicate that the pattern of weighting new, lower numbers more heavily is not specific to market research. The results do not support a purely motivational explanation for the pattern.

EXPERIMENT 5: A DIRECT TEST OF THE EVOKED-RANGE EXPLANATION

The previous experiments demonstrate that new, lower forecasts are overweighted in the integration process. Experiment 1A removes this effect and demonstrates that when subjects are provided an explicit range bounded by numbers close to the two projections, they integrate normatively. Both forecasts act as bounds to the range, and thus evoked range provides no information on the relative plausibility of each forecast. Although this result provides support for the assumed process that range information influences weights applied to the forecasts, more direct evidence is desirable. If the evoked range implies the plausibility of each forecast, it should be possible to create conditions in which the range implies that a second higher forecast is more plausible than the first. In such a situation, the results should be reversed such that movement from the first forecast is greater when the second forecast is higher (rather than lower) than the first. Experiment 5 accomplishes this goal by creating two conditions: one in which we provided participants with an equally skewed range (for which the lower and upper bound are equidistant from Projection A) and one in which we provided participants with an unequally positively skewed range (for which the upper bound is twice as distant from Projection A as the lower bound). In both conditions, the range is explicit, and participants know the range before Projections A and B; therefore, we expect that participants evaluate Projections A and B in the context of the explicit range. The range provides plausibility information, and thus, along with the margin of error, we expect that it affects the weights applied.

When the lower and upper endpoints are equidistant from Projection A, the evoked range is focused on the area of the distribution where the density is higher (frequency is more important than range; Parducci 1968). Neither explicit endpoint seems extreme, because both are equidistant from Projection A. When the second forecast Projection B is lower than Projection A, the density is concentrated below Projection A, and the range “lower bound to Projection A” seems more plausible. According to the frequency principle, the explicit upper endpoint is contrasted out, and the evoked range becomes the lower bound to Projection A. Similarly, when Projection B is higher than Projection A, the density is concentrated above Projection A, and the range "Projection A to upper bound" seems more plausible and becomes the evoked range. In both cases, Projection B is within the evoked range and Projection A is at the endpoint. Therefore, Projection A is considered less plausible than Projection B on the basis of the evoked range and is experimentally set up to be more plausible than Projection B on the basis of their respective margins of error. Thus, the two cues conflict both when Projection B is less than Projection A and when Projection B is greater than Projection A. We expect that the amount of movement from Projection A is the same in both conditions and that Projection B is likely to be weighted more heavily than is implied solely by margin of error in both the B-lower and B-higher conditions.

When one of the endpoints is much farther from Projection A than the other, as is the case in the unequally skewed range condition, both the density and the explicit range of the distribution affect the range evoked (both the range and the frequency principles are accorded weight; Parducci 1968). Thus, when Projection B is lower than Projection A, the range principle indicates that more weight should be applied to Projection A because it is closer to the center of the explicit range than is Projection B. However, the frequency principle indicates that evoked range should be concentrated on the region of the distribution where the density is higher (lower bound to Projection A), and the explicit upper endpoint should be contrasted out. This results in Projection B being applied more weight because it is contained in the relevant range. We expect that the compromise between the range and frequency principles results in a null effect of evoked range on integration; in other words, the evoked range is not perceived as informative about the relative plausibilities of Projections A and B. In this case, the only cue that drives the weights applied during integration is margin of error; thus, no bias is likely. When Projection B is higher than Projection A, there is a compromise again between the range and frequency principles, but in this case, both principles pull subjects toward Projection B. Thus, Projection B is judged as more plausible than Projection A on the basis of evoked range. Because Projection B is set up experimentally to be less plausible than Projection A (given its margin of error), the two cues provide conflicting evidence of relative plausibility. We expect that this results in Projection B being weighted more heavily than it should be, solely on the basis of its margin of error, which possibly results in a nonnormative estimate.

Method

Participants and design. Participants were 114 students who were paid $10 for their participation. The experiment used a 2×3 between-subjects design. The factors were Projection B being lower versus higher than Projection A and range (none versus equally skewed versus unequally positively skewed).

Stimuli and procedure. The stimuli (online grocery) and procedure were similar to those in Experiment 1. Participants in the no-range condition knew only the two projections and their margins of error. In the equally skewed and unequally skewed conditions, participants were given the projections, the margins of error, and the information about the minimum and maximum number of units the company could sell. The wording for the manipulation was as follows: "1000 is the minimum number of units your company could expect to sell annually and 7,465,000 (or 11,197,000, depending on the range condition) is the maximum number of units that your company could expect to sell annually." In addition, participants were asked to write down how they came up with their sales estimate immediately after they determined it.

Results

Sales estimates. An ANOVA on the amount of movement from Projection A revealed a significant interaction between type of Projection B (lower versus higher than Projection A) and range ($F(2, 106) = 4.61$, $p < .05$). The contrast between lower and higher Projection B in the no-range condition was significant (M lower = 1,144,421, M higher = 393,935; $F(1, 106) = 4.70$, $p < .05$). This replicated the result of previous studies. As we expected, the contrast between lower and higher Projection B in the equally skewed condition was not significant (M lower = 1,196,733, M higher = 934,308; $p > .4$). As we also expected, the contrast between lower and higher Projection B in the unequally positively skewed condition was significant and in the opposite direction (M lower = 688,850, M higher = 1,376,647; $F(1, 106) = 4.17$, $p < .05$). In this condition, there was greater integration of Projection B when it was higher (rather than lower) than Projection A.

Comparison to norm. In the no-range condition, the estimate participants provided was significantly lower than the norm when Projection B was lower (norm = 2,985,653, estimate = 2,588,579, $t[18] = -3.70$, $p < .005$) and at the norm when Projection B was higher (norm = 4,480,347, estimate = 4,126,935, $t[17] = -.98$). This replicates previous findings. In the equally skewed condition, the estimate provided was significantly lower than the norm when Projection B was lower (norm = 2,985,653, estimate = 2,536,267, $t[17] = -2.18$, $p < .025$) and significantly higher than the norm when Projection B was higher (norm = 4,480,347, estimate = 4,667,308, $t[17] = 2.03$, $p < .05$). In the unequally positively skewed condition, the estimate provided was at the norm when Projection B was lower (norm = 2,985,653, estimate = 3,044,150, $t[19] = .25$) and significantly higher than the norm when Projection B was higher (norm = 4,480,347, estimate = 5,109,647, $t[18] = 1.89$, $p < .05$). The results are fully consistent with the hypotheses we derived on the basis of range-frequency theory.

Open-ended responses. Participants indicated one or more of the following: (1) estimates somewhere between the projections given (n = 71, 62%), (2) taking into account of the margin-of-error rates of the research companies (n = 54, 47%), and (3) use of their own opinion (n = 25, 22%). Only five participants (4%) commented on the range, and all those participants were in the unequally positively skewed condition (three in the lower Projection B condition and two in the higher Projection B condition), which suggests that participants are not aware of evoking and using a range to judge plausibility when they naturally integrate two forecasts.

Discussion

Results provide strong support for the hypothesis that the evoked range influences how projections are integrated. Participants used the explicit sales ranges provided in their integration process by using them to judge the plausibility of the forecasts provided. This plausibility judgment, in turn, affected the weight applied during the integration process. Few participants even mentioned the range in their open-ended responses. It appears that participants are not consciously aware of the impact of ranges and do not evoke and use ranges intentionally.

To explore participants' awareness of their integration process further, we ran the no-range condition as Experiment 5A with two additions. Participants were asked to think aloud while they made their judgment. After making their judgment, participants were asked (1) whether they considered that fewer/more units might be sold than the report indicated and (2) what they considered the absolute minimum/maximum number of units that could be sold. Participants (n = 20) were executives with an average of 14 years' work experience.

Once again, the result replicates: More weight is given to Projection B when Projection B is lower (rather than higher) than Projection A (lower $w_B = .59$, higher $w_B = .13$). The concurrent verbal protocols made no reference to a range in which possible sales lie. However, in response to the specific questions, participants said that they were more likely to consider that fewer (rather than more) units than specified in the projections would be sold (within-subjects, 1 = "definitely no," 7 = "definitely yes"; fewer = 4.95, more = 3.48, $F(1, 19) = 7.46$, $p < .05$). This result supports the notion that participants believe that they are bounded by the highest number they have observed (in the absence of explicit information). Furthermore, 35% indicated that zero (or something close to zero) was the minimum number of units that could be sold. Only 14% indicated that infinity (or more than one million units) was the maximum.

GENERAL DISCUSSION

The results from five experiments and several follow-up studies consistently show that when decision makers are faced with the task of integrating discrepant forecasts, they use more information than only the associated margins of error. A series of experiments demonstrates that the additional information used is the range in which possible sales are expected to lie. This range is based on input such as previous sales. When no information about the range is available, decision makers appear to evoke a range from zero to the highest forecast of which they are aware. The range information is systematically used to infer the plausibility of the different forecasts, but its effects are not known to the decision maker. The influence of the range is so strong that it can even overwhelm the effect of forecast accuracy, as evidenced by margin of error. This can result in nonnormative forecasts (in which the norm is based solely on margins of error) and in greater weights being applied to less (rather than more) reliable forecasts. Overall, this article provides a robust demonstration that contextual information about the range in which forecasts can lie distorts the integration process by providing misguided input regarding the accuracy of forecasts.

We used range-frequency theory (Parducci 1965, 1995) to explain the influence of the evoked range. The plausibility of each forecast is influenced both by the range evoked and by the frequency of forecasts in different portions of the range. When the forecast is an endpoint of the range, it is viewed as more extreme and less plausible than a forecast that lies in the range. When there is a greater frequency of forecasts in a portion of the range, the forecasts seem less extreme and more plausible than forecasts in portions of the range in which the frequency is less dense. The final estimate is a joint function of these two principles.

This article's main contribution is its articulation of the influence of evoked ranges on judged plausibility and thus weight applied to different forecasts. It is critical for managers to understand the strong influence that evoked ranges can have on their decision making. In addition to research forecasts, managers often have access to both intuitive information about expected sales (what they believe is reasonable) and explicit information about the range of possible sales (e.g., prior sales, total market size). All this information affects which forecasts managers view as plausible and subsequently influences the final sales forecast. Although at times this may be appropriate, at other times this information can have a biasing influence on the forecast formed. For example, during the dot-com glory days, there were high expectations about sales across the marketplace. This "fever" of high sales may have served as input to the evoked range and caused people to weight higher forecasts more heavily. This research suggests that the range influences the weight applied to the forecasts and may overshadow the appropriate influence of the reliability of the forecasts. Whether a lower or higher second forecast is overweighted depends on where the forecasts lie in the range.

In some contexts, it may not be normative to focus only on the margin of error of the forecasts. The inclusion of plausibility judgments in the integration process may be considered appropriate if the plausibility is reasonably derived. In the experiments we report, the ranges evoked do not provide appropriate plausibility estimates. Market researchers are expected to consider the nature of the product, competition, and consumer response, whereas the evoked range ignores characteristics of the product market and consumer preferences. As such, evoked range provides misguided input about plausibility. Further research should explore when appropriate plausibility inferences are evoked or derived through an editing process. For example, new entrants may realize that they cannot steal share from the market leader, and thus they may evoke a sales range with market size less sales of the market leader as the upper bound. Although this editing process can lead to better intuitive forecasts, it still might not be appropriate to use it to complement objective market research estimates.

This article represents a start in the exploration of integration processes and their outcomes. Much work remains to be done to enhance the understanding of this. The most important question pertains to the source of the evoked range. What is the input that managers use to form ranges that are so consequential in their judgments? Would the range be normative (e.g., the two forecasts themselves) in some conditions, such as when managers produce the forecasts themselves? In addition, would plausibility inferences about the range affect the forecast produced if both projections were considered equally accurate? Other questions abound. For example, what happens if more than two forecasts are received; furthermore, what if the forecasts are all optimistic and suggest moving the lower bound upward from zero in the case of a new product and upward from previous year's sales for an existing product? What if the forecasts are received when sales have been declining over the years? What other contextual influences other than evoked range are used to determine plausibility and thus the weight given to different forecasts? In what conditions do motivational influences such as the negativity effect take center stage and supplant the perceptual distortion explicated here? Can managers be made aware of the influence of range and trained to correct for the effect? Finally, what are the conditions in which a range is not evoked to provide input to the integration process? For example, if the two forecasts are not as discrepant as the forecasts we have used, it is possible that extraneous range information will not be evoked and used. Given the well-acknowledged importance of forecasting in managerial decision making and the lack of research on how decision makers actually combine forecasts (Clemen 1989; Flores and White 1988), additional work is needed that addresses these issues.

REFERENCES

Armstrong, Scott J. (2001), *Principles of Forecasting: A Handbook for Researchers and Practioners*. Boston: Kluwer Academic Publishers.

Ashton, A.H. and R.H. Ashton (1985), "Aggregating Subjective Forecasts: Some Empirical Results," *Management Science*, 31 (12), 1499–1508.

Batchelor, Roy and Pami Dua (1995), "Forecasting Diversity and the Benefits of Combining Forecasts," *Management Science*, 41 (January), 68–75.

Bunn, Derek W. and Ahti A. Salo (1993), "Forecasting with Scenarios," *European Journal of Operational Research*, 68 (August), 291–303.

Clemen, Robert T. (1989), "Combining Forecasts: A Review and Annotated Bibliography," *International Journal of Forecasting*, 5 (4), 559–83.

Cooke, Alan D.J., Chris Janiszewski, Marcus V.M. da Cunha Jr., Suzanne Nasco, and Els DeWilde (2002), "Stimulus Context and the Formation of Consumer Ideals," working paper, University of Florida.

Cooper, Jim and John Consoli (1999), "USA Study Shows Sweeps Inflates Ad Rates," *Adweek*, 40 (December), 7.

Dalrymple, Douglas J. (1987), "Sales Forecasting Practices: Results from a United States Survey," *International Journal of Forecasting*, 3 (3–4), 379–92.

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

Dempsey, John. (1988), "After Grey Buyout, Debate Arises over LBS' Future," *Variety*, 333 (7), 48.

Devlin, Keith (2000), *The Math Gene: How Mathematical Thinking Evolved and Why Numbers Are Like Gossip*. New York: Basic Books.

Falconer, Robert T. and Charles M. Sivesind (1977), "Dealing with Conflicting Forecasts: The Eclectic Advantage," *Business Economics*, 12 (September), 5–11.

Figlewski, Stephen and Thomas Urich (1983), "Optimal Aggregation of Money Supply Forecasts: Accuracy, Profitability, and Market Efficiency," *Journal of Finance*, 38 (June), 695–710.

Fiske, Susan T. (1980), "Attention and Weight in Person Perception: The Impact of Negative and Extreme Behavior," *Journal of Personality and Social Psychology*, 38 (June), 889–906.

Flores, B.E. and E.M. White (1988), "A Framework for the Combination of Forecasts," *Journal of the Academy of Marketing Science*, 16 (Fall), 95–103.

Genest, Christian and James V. Zidek (1986), "Combining Probability Distributions: A Critique and an Annotated Bibliography," *Statistical Science*, 1 (1), 114–48.

Glazer, Rashi, Joel H. Steckel, and Russell S. Winer (1989), "The Formation of Key Marketing Variable Expectations and Their Impact on Firm Performance: Some Experimental Evidence," *Marketing Science*, 8 (Winter), 18–34.

Granger, C.W.J. and R. Ramanathan (1984), "The Improved Methods of Combining Forecasts," *Journal of Forecasting*, 3 (April–June), 197–204.

Green, Heather (1998), "The New Ratings Game," *BusinessWeek*, (April 27), 73.

Hogarth, Robin M. and Spyros Makridakis (1981), "The Value of Decision Making in a Complex Environment: An Experimental Approach," *Management Science*, 27 (January), 93–107.

Janis, I.L. and L. Mann (1977), *Decision Making*. New York: The Free Press.

Janiszewski, Chris and Donald R. Lichtenstein (1999), "A Range Theory Account of Price Perception," *Journal of Consumer Research*, 25 (March), 353–68.

Kahneman, Daniel and Dale T. Miller (1986), "Norm Theory: Comparing Reality to Its Alternatives," *Psychological Review*, 93 (April), 136–53.

Kang, H. (1986), "Unstable Weights in the Combination of Forecasts," *Management Science*, 32 (June), 683–95.

Lambert, D.R. (1980), "On Compensatory Demand Functions in Marketing Simulations," in *Experimental Learning Enters the Eighties*, D.C. Brenestuhl and W.D. Briggs, eds. Tempe, AZ: Bureau of Business and Economic Research, Arizona State University.

Lapide, Larry (2000), "New Developments in Business Forecasting," *The Journal of Business Forecasting Methods & Systems*, 19 (Spring), 16–18.

Larreche, Jean-Claude and Hubert Gatignon (1998), *MARKSTRAT3: The Strategic Marketing Simulation*. Cincinnati, OH: South-Western College Publishing.

Lichtenstein, Donald R. and William O. Bearden (1989), "Contextual Influences on Perceptions of Merchant-Supplied Reference Prices," *Journal of Consumer Research*, 16 (June), 55–66.

Lim, Rodney G. (1995), "A Range Frequency Explanation of Shifting Reference Points in Risky Decision Making," *Organizational Behavior and Human Decision Processes*, 63 (July), 6–20.

Makridakis, Spyros and Robert L. Winkler (1983), "Averages of Forecasts: Some Empirical Results," *Management Science*, 29 (September), 987–96.

Mentzer, J.T. and K.B. Kahn (1995), "Forecasting Technique Familiarity, Satisfaction, Usage, and Application," *Journal of Forecasting*, 14 (September), 465–76.

Morris, Peter A. (1977), "Combining Expert Judgments: A Bayesian Approach," *Management Science*, 23 (March), 679–93.

Niedrich, Ronald W., Subhash Sharma, and Douglas H. Wedell (2001), "Reference Price and Price Perceptions: A Comparison of Alternative Models," *Journal of Consumer Research*, 28 (December), 339–54.

Nunnally, Jum C. (1978), *Psychometric Theory*. New York: McGraw-Hill.

Parducci, Allen (1965), "Category Judgment: A Range-Frequency Model," *Psychological Review*, 72 (6), 407–418.

——— (1968), "The Relativism of Absolute Judgments," *Scientific American*, 219 (December), 84–90.

——— (1995), *Happiness, Pleasure, and Judgment: The Contextual Theory and Its Applications*. Mahwah, NJ: Lawrence Erlbaum Associates.

Sanders, N.R. and K.B. Manrodt (1994), "Forecasting Practices in U.S. Corporations: Survey Results," *Interfaces*, 24 (2), 92–100.

Schmittlein, David C., Jinho Kim, and Donald G. Morrison (1990), "Combining Forecasts: Operational Adjustments to Theoretically Optimal Rules," *Management Science*, 36 (September), 1044–56.

Sherif, M. and Carl I. Hovland (1961), *Social Judgment*. New Haven, CT: Yale University Press.

Singleton-Green, Brian (1990), "Auditors and the Expectation Gap," *Accountancy*, 105 (1160), 62.

Skowronski, John J. and Donal E. Carlston (1989), "Negativity and Extremity in Impression Formation: A Review of Explanations," *Psychological Bulletin*, 105 (January), 131–42.

Stone, M. (1961), "The Opinion Pool," *Annals of Mathematical Statistics*, 32 (December), 1339–42.

Telio, Steven (2001), Personal interview with the Marketing Manager, MyHelpDesk.com, (March 10).

Tetlock, Philip E. (1983), "Accountability and Complexity of Thought," *Journal of Personality and Social Psychology*, 45 (July), 74–83.

——— and R. Boettger (1989), "Accountability: A Social Magnifier of the Dilution Effect," *Journal of Personality and Social Psychology*, 57 (3), 388–98.

Urbany, Joel E., William O. Bearden, and Dan C. Weilbaker (1988), "The Effect of Plausible and Exaggerated Reference Prices on Consumer Perceptions and Price Search," *Journal of Consumer Research*, 15 (June), 95–110.

Volkmann, John (1951), "Scales of Judgment and Their Implications for Social Psychology," in *Social Psychology at the Crossroads*, John H. Rohrer and Muzafer Sherif, eds. New York: Harper, 272–96.

Wedell, Douglas H. and Allen Parducci (1988), "The Category Effect in Social Judgment: Experimental Ratings of Happiness," *Journal of Personality and Social Psychology*, 55 (September), 341–56.

Copyright of Journal of Marketing Research (JMR) is the property of American Marketing Association and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.