TANUKA GHOSHAL, ERIC YORKSTON, JOSEPH C. NUNES, and PETER BOATWRIGHT*

Marketers frequently offer a variety of communications, brands, and service encounters that customers evaluate sequentially. When customers make these evaluations, their previous experiences in the sequence influence their current evaluation. The authors propose that these prior experiences serve as multiple reference points against which the target stimulus is judged, creating rival co-occurring comparison effects. Using real-world and experimental data, they find that assimilation and contrast effects occur simultaneously: there is assimilation to the first score within a sequence and contrast with the immediate predecessor as well as with extremes experienced earlier in the sequence. The authors document the moderating effects of extreme first stimuli, domain similarity, and individual factors of mood and expertise. They provide different recommendations for sequence construction on the basis of whether the marketer's goal is fairness, accuracy, or influencing choice. This research is unique in (1) showing how several preceding evaluations can each have an impact on a subsequent evaluation at the same time and (2) using real-world data to do so.

Keywords: context effects, sequential judgments, hedonic evaluation, reference point effects

Online Supplement: http://dx.doi.org/10.1509/jmr.12.0075

Multiple Reference Points in Sequential Hedonic Evaluation: An Empirical Analysis

Marketers are often responsible for choosing the sequence for a series of hedonic events that consumers experience and evaluate one right after another. These sequences include similar items that are the object of intentional comparisons (e.g., wine tastings) as well as much different items that are likely to induce unintentional comparisons, such as the restaurant trend of "blind tastings." At 11 Madison Park in New York City, for example, customers can advise the chef about allergies, dietary restrictions, and ingredient preferences, but the composition of the daily $225 prix fixe menu is entirely up to the chef. Similarly, the custom of ordering *omakase* (Japanese for "I'll leave it to you") at sushi restaurants means that the customer lets the chef decide which dishes to prepare and in which order they are served. "Every course is relaxed and fun and exciting when the guest doesn't know what will come next," says Joseph Yaple, chef at the Four Seasons' restaurant One Forty in Lanai City, Hawaii (taste.fourseasons.com). The notion of an externally controlled sequence of experiences, in which the enjoyment of each episode may depend on what was previously encountered, is not limited to culinary events.

Managers and producers present audiences with various rosters of stand-up comics at comedy clubs and musicians at local bars, sometimes in what is presumably a random order (e.g., open mic night). Spectators often judge each performance either implicitly or explicitly, as is the case with most talent competitions. On the television reality singing competition *American Idol*, from the semifinal onward, the public calls in to vote during contestants' performances, which

*Tanuka Ghoshal is Assistant Professor of Marketing, Indian School of Business (e-mail: Tanuka_Ghoshal@isb.edu). Eric Yorkston is Associate Professor of Marketing, Neeley School of Business, Texas Christian University (e-mail: e.yorkston@tcu.edu). Joseph C. Nunes is Professor of Marketing, Marshall School of Business, University of Southern California (e-mail: jnunes@marshall.usc.edu). Peter Boatwright is Carnegie Bosch Professor of Marketing, Tepper School of Business, Carnegie Mellon University (e-mail: boatwright@cmu.edu). Direct correspondence to the first author. The authors thank the *JMR* review team for their direction throughout the project. Rik Pieters served as associate editor for this article.

ISSN: 0022-2437 (print), 1547-7193 (electronic)

air in a previously unspecified order. Competitors in other popular reality competitions—including *The Voice* (singing), *Project Runway* (fashion), *Design Star* (home design), *Face Off* (special effects makeup), *Top Chef* (cooking), and *America's Got Talent* (variety)—are evaluated similarly by either the public or the show's judges. Understanding how earlier events influence evaluations of subsequent events is critical to achieve efficient (i.e., accurate) and fair outcomes when ordering the sequence (Page and Page 2010).

The outcomes of sequential evaluations matter to marketers in many other contexts as well. Customer feedback is used to help improve future experiences and is often encouraged in the form of sequential evaluations. The free Internet radio station Pandora, for example, provides a random mix of songs that listeners can evaluate by choosing "thumbs up," "thumbs down," or "never play." On Pandora, listeners do not choose which songs they hear or the order in which they hear them; instead, their responses are intended to help fine-tune playlist preferences, thereby improving the chance that users will hear more of what they like in the future. In the video game industry, players are known to judge each level or "map," and their evaluations influence subsequent reviews on fan sites, which in turn affect demand. Thus, future sales depend on game developers creating "design patterns" that enhance each stage of the gamer's experience (Bjork and Holopainen 2004). Similarly, filmgoers' evaluations of various movie trailers influence which movie they decide to see in the future. With this knowledge, theater and studio executives deliberate on which trailers will be shown with a particular film and in what order.

Our research focuses on these types of consumption episodes. We document the influence of multiple reference points within a particular sequence of hedonic experiences with the expectation that earlier events within the sequence serve as reference points against which later events are simultaneously compared. Thus, this research presents the first empirical test of co-occurring and potentially opposing comparative effects by consumers who evaluate experiences in sequence. In line with our expectations, our data reveal the impact of multiple reference points, with both kinds of comparative effects—assimilation and contrast—simultaneously influencing a single evaluation. Using both real-world and experimental data, we find a similar pattern of effects: assimilation to the first event in the sequence, contrast with the immediately preceding event, and contrast with the best and worst prior event in the sequence.

Our first data set comprises eight years of evaluations from a national home-brewing competition in which hundreds of trained judges tasted thousands of beers. We observe that scores on subsequent beers are influenced by judgments of more than one prior beer within a sequence in predictable ways. These effects are enhanced when there is greater similarity among the beers within a particular sequence. We replicate the results of our beer data with a controlled experiment conducted online, which helps rule out the impact of extraneous variables and provides further evidence of the reliability of our results. It is reassuring that we observe the same pattern of effects in the real world as we did in the lab (Brinberg and McGrath 1985; Lynch et al. 2012); in doing so, we address the issue of "realism" raised by researchers concerned with the field's trend toward "methods myopia" (Davis et al. 2013; Hubbard and Lindsay 2002). These kinds of empirical findings are needed in marketing (Deighton et al. 2010; Ehrenberg 1995; Lynch 2011).

In our online experiment, several hundred respondents evaluated sequences of 20 jokes. Our choice of stimuli bolsters the generalizability of the effects by extending our findings from a primarily sensory experience (taste) to a primarily cerebral one (humor) and away from a task in which the goal is to choose the best option. The experimental design also enables us to test the robustness of our results in the face of (1) an extremely positive or negative first experience; (2) varying levels of expertise among judges; and (3) variations in mood, which are a potential influence on evaluative judgments (for a review, see Schwarz and Clore 1996).

By documenting how evaluations of prior experiences can serve as reference points with co-occurring opposing effects, our findings contribute to several streams of research. First, we extend the research on sequential evaluation by showing a "sequential history bias" (Page and Shapiro 2010, p. 188) to multiple reference points simultaneously. A sequential order bias arises when the target of the evaluation is advantaged or disadvantaged by its position in the sequence. Previous work on order effects in sequences has focused on comparisons step by step, or how a new option relates only to the one immediately preceding it (Bruine de Bruin and Keren 2003). Our research is unique in showing how several preceding evaluations can each affect a subsequent evaluation at the same time.

Second, by investigating the impact of earlier evaluations as reference points, we extend the research on comparative effects. Judgments of a target stimulus have been shown to assimilate to a previously judged stimulus (Damisch, Mussweiler, and Plessner 2006) or a previously encountered anchor (see, e.g., Chapman and Johnson 1994). Other research has shown that judgments of moderate stimuli are contrasted with more extreme stimuli (Herr 1986; Wyer and Srull 1989). Mussweiler and Strack (2000) argue that assimilation and contrast are not necessarily exclusive processes, showing that the same reference point can produce assimilation effects for objective evaluations and contrast effects for subjective evaluations. Notably, their study involves two separate judgments derived from a single reference point. Our research differs in that we demonstrate how a single evaluation can simultaneously assimilate toward one reference point while contrasting away from other reference points.

Third, this article helps connect prior research on hedonic evaluations with research on multiple reference points. The extant research on multiple reference points has been limited to studying the influence of goals and similar types of benchmarks as standards. We build on prior research in this area by using a person's own prior judgments as the multiple standards of comparison. Fourth, we also attempt to resolve some of the conflicting findings in the literature on contrast and assimilation effects regarding the issue of domain match. Fifth, by documenting how extreme initial experiences, expertise, reference similarity, and mood interact with context effects, our results open the door to many avenues for further research. The results presented herein, both from the lab and the real world, should encourage new theorizing regarding when, how, and why multiple prior

judgments serve as reference points (Lehmann, McAlister, and Staelin 2011; Reiss 2011).

In summary, the data and our analyses enable us to (1) identify an important phenomenon—the simultaneous influence of multiple reference points in hedonic evaluation—(2) introduce compelling evidence for the validity of the effects, and (3) highlight several important moderators that should help guide further research. We organize the remainder of this article as follows. First, we briefly review the relevant research pertaining to multiple reference points, comparative effects in sequential evaluation, and hedonic evaluation. Next, we describe the beer tasting data set that enables us to test for and uncover the presence of co-occurring assimilation and contrast effects in the real world. We then describe the online experiment that demonstrates the robustness of the effects found in the beer data while testing key moderators including expertise, mood, reference similarity, and extreme initial experiences. We conclude with a discussion of the implications of our findings for managers and further research.

BACKGROUND

Multiple Reference Points

It is generally accepted that people's evaluation of a stimulus depends not only on the experience itself but also on context (Dato-on and Dahlstrom 2003; Podsakoff et al. 2003). Context effects often include a single reference point but can include multiple relevant reference points during the decision-making process. When confronted with multiple reference points, scholars have taken different approaches to model their influence. For example, early pricing research assumed that people combine multiple reference points into a single standard of comparison, such as when a reference price is based on the average of prices observed in the past (Winer 1986). Subsequent work by Mayhew and Winer (1992) demonstrates that there can be joint effects from more than one reference point, such as when internal and external reference prices each have an influence on purchase decisions. Other research has shown that people can keep multiple reference points (e.g., salaries) separate in their minds and that these different standards can have varying effects depending on the task at hand—including, for example, whether a person is assessing satisfaction as opposed to assessing fairness (Ordóñez, Connelly, and Coughlin 2000). Similarly, March and Shapira (1992) show how people might attend to both the minimum required (survival point) and the goal (aspiration level) as multiple reference points but assume that people draw on either one or the other exclusively when making a decision.

More closely related to our research is "tri-reference point theory," which considers the effect of three standards as reference points within the same choice context: (1) the minimum requirement, (2) the status quo, and (3) the goal or aspiration level (Wang 2008). Koop and Johnson (2012) suggest that these three standards can each have a differential yet joint impact on choice behavior by showing how people's choices among a set of gambles demonstrate an awareness of the minimum requirement, the status quo, and the goal. However, previous research has been unable to distinguish whether respondents in the experiments are demonstrating an adaptive use of the differing reference points or consider all the reference points throughout their judgment process. Koop and Johnson (2012, p. 60) emphasize that they cannot "draw strong conclusions about the simultaneous use of the reference points." In this article, we build on previous research by presenting direct evidence of the simultaneous use of multiple reference points. We also show how multiple reference points affect a series of judgments rather than merely a single choice. Furthermore, with our unique data sets, we answer Koop and Johnson's (2012, p. 60) call for research using "less artificial" reference points in more "lifelike contexts" to "affirm the increase in descriptive power of multiple-reference-point dependence as a model for real-world decision scenarios."

Comparative Effects in Sequential Evaluation

Within a sequence of evaluations, any single evaluation after the first can be influenced by a prior evaluation (Ariely, Loewenstein, and Prelec 2003). Consequently, the effects of comparison processes often appear in sequential judgments. Damisch, Mussweiler, and Plessner (2006) demonstrate this phenomenon when examining judgments between pairs of athletes appearing successively in sporting events. They find evidence that judgments of a target performance are influenced by the previously judged performance, and the direction of this influence (assimilation or contrast) depends on the degree of perceived similarity between successive performances. This pattern is consistent with Mussweiler's (2003) selective accessibility model, which allows for two outcomes depending on which type of fundamental comparison process takes place. In his study, when participants considered the perceived similarity between the target and the standard (similarity testing), the result was assimilation (i.e., a positive relationship between the standard and the target). Conversely, when they searched for information regarding how the standard and target were dissimilar (dissimilarity testing), the result was contrast (i.e., a negative relationship between the standard and the target). It is worthwhile to point out that there are some conflicting findings with respect to this issue in the marketing and psychology literature streams. Whereas Schwarz and Bless (1992) similarly find that inclusion within the same mental category (similarity) leads to assimilation and exclusion (dissimilarity) leads to contrast, several other researchers have found that similarity enhances contrast and dissimilarity enhances assimilation (Raghunathan and Irwin 2001; Zellner, Kern, and Parker 2002). In the present work, we attempt to shed some light on why these conflicting results may arise. Our real-world data account for stimuli similarity because the beers are classified into categories and subcategories such that beers within a subcategory are more similar to one another, enabling us to assess the impact of subcategory changes within a sequence.

The role of information accessibility on comparative processes rests on the assumption that people choose a single standard of comparison (Mussweiler 2003), which has historically been the preceding stimulus in research on sequential evaluation. We challenge this assumption by proposing that people often use multiple standards of comparison simultaneously, including more than one predecessor. In Damisch, Mussweiler, and Plessner (2006), as in all of the assimilation–contrast research we could identify in psychol-

ogy and marketing, the researchers focus on pairwise comparisons (i.e., testing the relationship between the target stimulus and a single standard). We believe that what has been found to occur during the assessment of two consecutive stimuli is likely to occur during the assessment of *multiple* consecutive stimuli; each target can be compared with multiple prior stimuli. Unlike previous research in sequential evaluation, we are the first to test for and document the co-occurrence of comparative effects to multiple reference points within a sequence.

Research on psychophysical judgments, social comparisons, and personal perceptions has suggested that it is difficult to predict a priori the type of comparative effects that will occur (Damisch, Mussweiler, and Plessner 2006). In anticipating which experiences might serve as reference points, we turned to prior research for guidance. Although extensive literature exists on order effects, it still is not possible to predict with any accuracy whether the first or last (most recent) experience within a sequence will have a significant effect (Kardes and Herr 1990). Consequently, we test for both the first and last (i.e., the stimulus experienced immediately preceding the target) experiences in the sequence. Previous research has also found that contrasts occur for comparisons with extreme standards (Herr 1986; Herr, Sherman, and Fazio 1983). Thus, we test for comparative effects to the highest and lowest prior evaluation within the sequence at the time the target is being evaluated. These extreme points can be considered two endpoints of the stimulus range (Parducci 1964).

Hedonic Evaluation

The current research examines the impact of multiple reference points on evaluations made within a sequence of hedonic experiences. Previous research investigating assimilation and contrast in hedonic evaluations within a sequence has focused exclusively on how one experience influences the next experience and has done so exclusively in the lab. This type of procedure and analysis has produced mixed results. For example, Novemsky and Ratner (2003) find that although people tasting pairs of jelly beans expect contrast effects beforehand and recall contrast effects afterward, evaluations provided at the time of the actual experience provide no evidence of contrast effects. Conversely, Zellner, Kern, and Parker (2002) find contrast effects in real-time sequential evaluations of pairs of fruit juices of differing dilutions. Like Zellner et al. (2003), and contrary to Novemsky and Ratner (2003), we find contrast effects in real-time sequential hedonic evaluations. Unlike Zellner et al. (2003), we also test for and find simultaneously co-occurring assimilation effects.

We extend prior research on hedonic evaluation by investigating sequences of more than two stimuli and testing for the effects of multiple points of comparison. Furthermore, we do so using both real-world data and experiments. By *not* focusing only on pairwise comparisons of successive stimuli, our analysis enables us to document simultaneously occurring assimilation and contrast effects. By showing how controlling for one type of comparative effect to one reference point can reveal the presence of another comparative effect to a different reference point, our work may shed light on why some researchers observe contrast effects whereas others find assimilation effects, and still others find no effects at all.

REAL-WORLD DATA ANALYSIS: BEER TASTING

Data Description

We obtained eight years of judging data from the Bluebonnet Brew-Off, a national brewing competition held annually in the Dallas/Fort Worth Metroplex. From 2000 to 2007, more than 900 brewers entered more than 6,000 beers into competition in 23 style categories (comprising 107 subcategories). Organizers randomly assigned beers from the same category into tasting "flights" ranging from 5 to 13 beers. There are 688 flights in our data. Beers within a flight belonging to the same subcategory form a "subflight." For example, brown porter, robust porter, and Baltic porter are all subcategories of the porter category. Organizers randomly ordered beers within each subflight, with the only constraint being that beer subcategories be ordered from light to dark within a flight.

Two, and occasionally three, independent judges (average correlation between judges = .92) rated each beer in a flight in the same order. The data include the judges' summary score (out of 50 points) for every beer evaluated. Each score was recorded while tasting the specific target beer and written down before moving on to evaluate the next beer; thus, our data comprise sequences of real-time hedonic evaluations. The top beers in each flight moved into a second round, in which they were evaluated again in category flights designed in a fashion similar to the first round. The highest-ranking beers in the second round of scoring were declared the winners of their respective categories.

This data set is ideal for the purpose of our research objective because it contains multiple instantiations of sequential evaluation processes in random sequences (688 flights evaluated by two or three judges, resulting in 1,423 unique sequences of 5–13 beers). Finding context effects in these data would speak to the pervasiveness and strength of these types of effects because the judges were trained to be observant for signs of satiation and fatigue (Strong and Piatz 2008). The Web Appendix contains additional details regarding the data as well as preliminary analyses, which include testing for signs of positive and negative sequential drift used to rule out any role satiation, preference for darker beers, or indications that palate fatigue may have played a role in driving our results.

Model

We conducted all analyses at the flight level for an individual judge. To test the simultaneous effect of the preceding beer and potential extremes (i.e., the running maximum and minimum scores assigned to previous beers in the flight), we estimated coefficients only for beers in the third position and later because these variables are not identified for the first two beers in the sequence. We used a hierarchical Bayesian model to allow for heterogeneity across flights because we cannot assume that (1) each parameter is identical across flights or (2) each flight is unrelated and independent of other flights. The model estimates coefficients considering beer k, $k \in (3 \ldots 10)$, in flight i, $i \in (1 \ldots 688)$, evaluated by judge j, $j \in (1 \ldots 3)$, across the eight years (2000–2007). The resulting model is as follows:

$$(1)\ y_{ijk} = \beta_{0ij} + \beta_{1ij}y_{ij,k-1} + \beta_{2ij}\text{FIRST}_{ij} + \beta_{3ij}\text{MAX}_{ijk} + \beta_{4ij}\text{MIN}_{ijk} + \beta_{5ij}\text{SUBCATCHANGE}_{ik} + \beta_{6ij}\text{SUBCATCHANGE}_{ik} \times y_{ij,k-1} + \beta_{7ij}\text{SUBCATCHANGE}_{ik} \times \text{MAX}_{ijk} + \beta_{8ij}\text{SUBCATCHANGE}_{ik} \times \text{MIN}_{ijk} + \beta_{9ij}\text{POSITION}_{ik} + \beta_{10ij}\text{ROUND}_{ik} + \beta_{11ij}\text{YEAR}_{ik} + \beta_{12ij}\text{STYLE}_{ik} + \varepsilon_{ijk},$$

where y_{ijk} represents the score (out of 50) of the kth beer in the ith flight as assigned by the jth judge. We include the score of the first beer in the ith flight as assigned by the jth judge directly as an independent variable (FIRST_{ij}). The immediate prior beer is represented in the model by y_{ijk-1}; MAX_{ijk} is the score of the running maximum beer up to the focal beer y_{ijk}; MIN_{ijk} is the running minimum; SUBCATCHANGE_{ik} is a dummy variable indicating whether the focal beer belongs to a different beer subcategory from the previous beer; and YEAR_{ik}, ROUND_{ik}, POSITION_{ik}, and STYLE_{ik} are vectors of dummy variables to control for potential effects of year, round, position, and beer style, respectively. Note that significant positive coefficients reflect positive correlations, indicating assimilation effects, whereas significant negative coefficients reflect negative correlations, indicating contrast effects. To illustrate, if the sign on the parameter for score of the previous beer β_{1ij} is negative, it means that the score of the focal beer contrasts with this value, and a higher (lower) score of the previous beer would indicate a lower (higher) score for the focal beer.

The vector of coefficients β_{ij} represents judge–flight-level random effects, whereas the intercept term β_{0ij} would be judge–flight-level intercepts, representing flight-level assignment effects. Following Rossi, Allenby, and McCulloch (2005), we specified the priors on the collection of β_{ij} using a two-stage process. First, we specify a normal prior (with a common mean vector $\overline{\beta}$ and a fixed variance matrix V_β) on β_{ij} (Equation 2), and then we specify a second-stage prior on the parameters of this distribution.

$$(2)\quad \beta_{ij} \sim N(, V_\beta)$$
$$\overline{\beta} \sim N(0, A^{-1})$$
$$V_\beta \sim IW(\nu, V).$$

The prior and posterior for the covariance matrix of the multivariate normal distribution are of the inverted Wishart form; V_β represents the unobserved heterogeneity, with V characterizing the extent of the heterogeneity; all notations and characterizations follow Rossi, Allenby, and McCulloch (2005). The second-stage priors are set to be proper but diffuse ($A = .01I$; $\nu_{ij} = 3$; $\nu = k$ [number of variables] + 3; $V = \nu \times I_k$). Because we have a large number of units in the analysis, we expect the data to overwhelm these priors. Inferences about specific flights may be obtained from the posterior distribution of β_{ij}.

The algorithm for estimating the model parameters is a Markov chain Monte Carlo (a Gibbs sampler) that draws recursively from the posterior conditional distribution of the model parameters. We followed the model estimation procedure of Allenby and Rossi (1999), Gelfand and Smith (1990), and Rossi, Allenby, and McCulloch (2005). We used 100,000 draws to estimate the model with initial parameter values set to 0. Time-series plots of the draws indicate that the distributions converged to stationary distributions; we confirmed convergence using the Heidelberger and Welch convergence diagnostic as implemented by the Bayesian output analysis program in R. Table 1 presents the model parameter estimates.

ROUND, the variable representing whether the beer was evaluated in round 1 or round 2, was significant and positive (posterior mean ROUND = 2.53, $p < .05$). Beer scores in the second round consisted only of winning beers from the first round and, on average, were higher. The variable controlling for YEAR was not significant, nor was POSITION, meaning that there were no observed position effects for beers later in the flight due to preference for lighter or darker beers, taste satiation, or judging fatigue. The variable STYLE indicated that there were significant differences in

Table 1
MODEL PARAMETER ESTIMATES FOR BEER DATA

Parameter	*Variable*	*Variable Description*	*99% Posterior Interval*[a]	*Posterior Estimate (SD)*
β_{0ij}	Intercept	Intercept	(4.99, 54.6)	46.2 (3.5)***
β_{1ij}	y_{ijk-1}	Score of previous beer	(–.13, –.03)	–.08 (.02)***
β_{2ij}	FIRST	Score of first beer in flight	(.37, 1.58)	.64 (.11)***
β_{3ij}	MAX	Score of running maximum	(–.69, –.42)	–.56 (.05)***
β_{4ij}	MIN	Score of running minimum	(–.67, –.40)	–.55 (.05)***
β_{5ij}	SUBCATCHANGE	Dummy = 1 if new subcategory within flight, and 0 otherwise	(–16.2, 4.12)	–8.29 (3.83)*
β_{6ij}	$y_{ij,k-1}$ × SUBCATCHANGE	Interaction of subcategory dummy and score of previous beer	(–.04, .27)	.11 (.06)**
β_{7ij}	MAX × SUBCATCHANGE	Interaction of subcategory dummy and score of running maximum	(–.15, .35)	.12 (.09)
β_{8ij}	MIN × SUBCATCHANGE	Interaction of subcategory dummy and score of running minimum	(–.29, .27)	–.008 (.11)
β_{9ij}	ROUND	Round (1 or 2)	(–.25, 6.77)	2.52 (1.23)**
β_{10ij}	YEAR	Year	(–.41, 1.06)	.28 (.28)
β_{11ij}	POSITION	Position in flight	(–.24, .09)	–.08 (.06)

*$p < .10$.
**$p < .05$.
***$p < .01$.
[a]Interpreted as the posterior probability that the parameter β_{0i} lies in the interval 4.99 to 54.6 is .99.

scores per the subcategory of beer. Certain styles of beer, such as American lagers and porters, received higher scores, on average, whereas subcategories such as light 60 Scottish ales received lower scores. Organizers stated that the list of higher-scoring subcategories was consistent with their belief that certain beer styles are more difficult to produce, and only more proficient brewers tended to enter beers in those subcategories. The Web Appendix includes a table of the significant dummy variables.

Results

Assimilation. The marginal posterior for the score of the first beer is positive (posterior mean FIRST = .64, $p < .01$), signifying an assimilation effect. The higher (lower) the score of the first beer, the higher (lower) the score of the focal beer. In particular, it seems that the rating of the first beer acts as an anchor (Ariely, Loewenstein, and Prelec 2003; Chapman and Johnson 1999; Tversky and Kahneman 1974) and thus serves as a point of reference (Lynch, Chakravarti, and Mitra 1991). In addition, our preliminary analysis revealed that the mean of the ratings of beers in the first position of the sequence is significantly higher than the mean of the ratings of beers in subsequent positions ($M_{position1} = 33.5$, $M_{rest} = 32.4$; $p < .01$). This finding parallels research showing a primacy bias in food such that the first item sampled is experienced most strongly and is more memorable (MacFie et al. 1989; Welch and Swift 1992).

Contrast. Analysis of the marginal posteriors for the score of the previous beer (posterior mean $y_{ij,k-1} = -.08$, $p < .01$), the score of the running maximum (posterior mean MAX = –.56, $p < .01$), and the score of the running minimum (posterior mean MIN = –.54, $p < .01$) reveals that more than 99% of the mass is below zero. The negative correlation between the focal beer and the previous beer indicates contrast effects between adjacent beers. The negative correlation with the running maximum and minimum scores indicates contrast effects between the target beer and previous extreme scores.

When evaluating beers, judges implicitly compared each beer tasted with exemplars of the beer style, with the ultimate goal of selecting a category winner. Consistent with previous research that has investigated extreme exemplars (Herr 1986; Herr, Sherman, and Fazio 1983), we find contrast effects for prior beers that are especially good or especially bad (i.e., extremes). Taken together, these results document how the evaluation of a focal beer is influenced by assimilating to the evaluation of the first beer while simultaneously contrasting with the evaluation of the beer immediately preceding it as well as with preceding extreme (good and bad) scores.

Similarity effects. All beers within a specific flight are similar in that they belong to the same category (e.g., porter). However, within a category (flight), there are frequently beers from two or more subcategories (e.g., brown porter, robust porter, Baltic porter) that may be regarded as less similar. For the first beer of a new subcategory, the preceding beer is the last beer of the previous subcategory and could be classified as a stimulus from a "different domain." The interaction term of the subcategory change dummy variable with the score of the previous beer was significant and positive (posterior mean $y_{ij,k-1} \times$ SUBCATCHANGE = .12, $p < .05$). Given that the distribution on the scores of the previous beer is negative (indicating a contrast effect), the positive sign on the interaction term implies that a category change attenuates this effect. In other words, when two adjacent beers in a sequence belong to different subcategories, the contrast effect is weaker. Intuitively, if the subcategories are different, the extent to which the previous beer is perceived as a useful standard should be reduced. Conversely, if two adjacent beers are from the same subcategory, the contrast effect between the target beer and the previous beer should be and is found to be stronger. This result is consistent with Redden (2008), who finds that subcategorization focuses attention on differentiating aspects, and with Rota and Zellner (2007), who find that experts (such as the judges in our data) categorize at a finer level and would contrast less between stimuli belonging to different categories. Finally, our results are consistent with prior literature (Raghunathan and Irwin 2001; Zellner, Kern, and Parker 2002) that has shown enhanced contrast with enhanced similarity, but they are at odds with other research (Schwarz and Bless 1992). We believe the disparity in findings may be because we (and the former set of researchers) study hedonic or subjective evaluations, whereas Schwarz and Bless (1992) study objective judgments.

The marginal posterior for the subcategory change dummy variable is marginally significant and negative (posterior mean SUBCATCHANGE = –8.29, $p = .07$), indicating that the score of the first beer in a new subflight is adjusted downward. Although we had no reason to expect this effect a priori, it is possible that judges used this opportunity to recalibrate their scales (Lynch, Chakravarti, and Mitra 1991) to avoid encountering a ceiling effect. The interaction terms of the subcategory change dummy variable with the scores of the running maximum and minimum beers were not significant.

Discussion

The estimation reveals multiple sources of influence (i.e., the systematic influence of multiple reference points) on the evaluation or score of a single beer within a flight. For a given beer, there are assimilation effects to the score of the first beer and contrast effects with the previous beer, as well as with prior extremes (running maximum and minimum scores). Our model has enabled us to test for and confirm the influence of co-occurring assimilation and contrast effects on the assessment of individual target evaluations within a sequence of evaluations and assess the differential magnitude of their impact.

It is important to recognize that the assimilation and contrast effects we uncover had substantively significant as well as statistically significant influences on evaluations. For example, consider the posterior estimate for the running maximum ($\beta_{3ij} = -.56$). Translating the magnitude of this estimate in terms of scores, an increase of two standard deviations (equivalent to 12.6) in the score of the running maximum (which ranges from 15 to 50) would result in an average impact of –7.0 (a seven-point decrease) on the score of the focal beer. In our data, 33% of winning beers in a flight are decided by one point or less, and less than a two-point difference decides 58% of final category winners. These findings mean that a highly rated beer early in the flight can keep subsequent beers out of medal contention. The contrast and assimilation effects induced by the running

minimum, score of the previous beer, and score of the first beer are of similar magnitudes in terms of their impact. In Table 2, we compute effect sizes for one standard deviation change in these variables on the dependent variable. It seems that randomizing the order of experiences does not solve the problem of context effects. We show that context effects can have a profound impact on evaluative outcomes even in randomized sequences.

ONLINE EXPERIMENT: ENJOYMENT OF JOKES

When designing the online experiment, we focused on several objectives. First, we set out to replicate the results from our analysis of the beer data in a different domain. Second, our analysis of the beer data relied on real-world data. Although these data ensured a high degree of external validity, they prevented us from answering several relevant questions pertaining to our results. For example, we found assimilation to the first stimulus, and yet all beers were relatively high in quality and were ordered randomly. Our beer data did not provide the opportunity to assess how the results might have differed if the first experience were extremely positive or negative. In addition, all the beer judges possessed a certain amount domain-specific knowledge. We believe that it is important to assess the role that expertise might have played, if any, in exacerbating or attenuating the effects we observed. We designed an online experiment to test the robustness of our findings while accounting for differences in relative expertise and the valence and extremeness of the first stimulus evaluated in the sequence.

Designing an experiment also gave us the opportunity to assess and control for the role of mood during sequences of real-time hedonic evaluations. Raghunathan and Irwin (2001) find evidence for mood-based assimilation for recalled experiences; when enjoyment of the stimuli put participants in a positive mood, assimilation to the target was observed. Because tasting beer and evaluating jokes may cause a similar type of mood effect, we include mood measures as a potential moderator in this study.

Data Description

Pretest. We gathered 100 jokes from joke books and online joke collections to create our initial set. In a pretest, 354 U.S.-based Amazon.com Mechanical Turk (MTurk) users evaluated one of four sets of 25 jokes, rating each joke on a seven-point enjoyment interval scale (1 = "did not enjoy at all," and 7 = "enjoyed a lot"). From these evaluations, we selected 20 jokes for the main study that varied in overall enjoyment but were consistent in how they were rated across judges (i.e., these jokes had a range of means but low standard deviations). In addition, we selected two jokes that were uniformly disliked and well liked, respectively, to be used as the "extremely negative" and "extremely positive" stimuli ($M_{negative} = 2.04$, SD = 1.55 vs. $M_{set} = 4.07$, SD = 1.11; $t(91) = 7.6$, $p < .01$; $M_{positive} = 5.63$, SD = 1.36 vs. $M_{set} = 4.22$, SD = .89; $t(91) = 6.38$, $p < .01$). The Appendix presents the extremely negative and extremely positive jokes. The full text of the 20 jokes used in the study and all 100 pretested jokes are available from the authors upon request.

Main study. We ran the main study online using participants from three sources: MTurk with U.S. citizens (N = 400), student participants from a major U.S. university (N = 143), and students and staff from an English-speaking international university (N = 212). Participants completed the task for either cash incentives or course credit. We randomly assigned each participant to one of three conditions in a three-level single factorial design: (1) first stimulus extremely negative, (2) first stimulus extremely positive, or (3) first stimulus randomly assigned from the 20 jokes (control condition). For conditions 1 and 2, the rest of the jokes were presented in random order, whereas for the control condition, all 20 jokes (including the extreme ones) were presented in random order. In other words, only the first joke was fixed to be extremely negative or positive in the first two conditions; otherwise, each respondent read and responded to a sequence of jokes ordered randomly.

Participants were told that they would be asked to complete a survey designed to assess how jokes may be more or less appealing to different people. After signing an online consent form, they indicated their current mood on a seven-point scale anchored from "very negative" to "very positive," because mood can affect joke enjoyment (Raghunathan and Irwin 2001; Schwarz and Clore 1983). Participants then evaluated the 20-joke set using the seven-point enjoyment scale from our pretest. Next, respondents completed Kleiser and Mantel's (1994) self-rated consumer expertise scale adapted for the humor domain (available in the Web Appendix). We were comfortable relying on self-reports of expertise because Wright et al. (1994) suggest that, in general, self-rated expertise is a good predictor of a person's relative performance. We also collected some objective measures for expertise in terms of familiarity and knowledge of different categories of jokes as well as some behavioral measures regarding entertainment preferences in the humor genre. Our intention was to assess whether participants' self-reported expertise correlated with our more objective measures (also provided in the Web Appendix). Finally, after participants provided their age and gender, they were thanked for their participation and debriefed.

Table 2
CONTEXT EFFECT SIZES: BEER DATA

Variable	*Range in Data*	*Coefficient; Model 1*	*Impact on Score of Focal Beer for 1 SD Increase*
Score of previous beer	12–49.5	–.08	Average impact –.5 decrease in score
Score of first beer	14–47.5	.64	Average impact +4.0 increase in score
Score of running maximum	17.5–49.5	–.56	Average impact –3.5 decrease in score
Score of running minimum	12–42	–.55	Average impact –3.5 decrease in score

Notes: 1 SD in beer scores = 6.3. Significance of effects: 33% of winners in the Brew-Off are determined by a margin of one point or less; 58% of winners are determined by a margin of two points or less.

Model

Preliminary analysis. Of the 755 records, 5 had been erroneously duplicated and were deleted from the data. Thirty-one participants failed the attention check question (see the Web Appendix), with 10 completing the survey in less than 3 minutes (average time for completion was 14 minutes). In addition, these 10 participants provided the same response across questions, indicating that they had mindlessly clicked through the questions, and were therefore removed from the study. The remaining 21 participants who had failed the attention check question were retained in the data; the results did not substantively change with or without their inclusion. The final sample included responses from 740 respondents (41% female; average age = 29.5 years) from the three sources as follows: MTurk with U.S. citizens (N = 398; average age = 32.5 years; 44% female), student participants from a major U.S. university (N = 136; average age = 20.7 years; 49% female), and students and staff from an English-speaking international university (N = 206; average age = 29.3 years; 32% female). Descriptive statistics (e.g., average rating, expertise) reveal no significant differences between the three sources (see the Web Appendix); however, because of the differences in the origin and composition of the three samples, we account for participant source in our model specification of heterogeneity.

The average rating for jokes in our data was 4.29 on a seven-point scale, with a standard deviation of .9. As a manipulation check, we compared the average rating for the joke that was preselected to be extremely negative in the first position with the average for randomly assigned first jokes in the control condition. Participants rated the extremely negative joke significantly lower ($M_{firstneg}$ = 2.32, SD = 1.45; $M_{firstcontrol}$ = 4.17, SD = 1.61; t(416) = –12.9, $p < .01$). Similarly, the average rating for the joke that was preselected to be extremely positive was significantly higher than the average first joke in the control condition ($M_{firstpos}$ = 5.48, SD = 1.35; $M_{firstcontrol}$ = 4.17, SD = 1.61; t(398) = 9.37, $p < .01$). These results confirm that our manipulation of extremes for the first joke in the sequence was effective. Self-rated expertise in the sample averaged 4.4 (SD = 1.10; minimum = 1.0, maximum = 7.0) and was positively correlated with the objective measures of expertise (r = .43, $p < .05$).

Model and estimation. Our data comprise 740 sequences of 20 evaluations. Given the multilevel nature of the data, and to allow for the most flexibility in the error structure across levels, we formulate and estimate a flexible, hierarchical, linear Bayesian random parameters model as we did with our beer data. The model consists of a set of regression equations with different error variances for each equation, $y_i = X_i\beta_i + \varepsilon_i$; $\varepsilon_i \sim$ i.i.d. $N(0, \sigma_i^2)$, allowing for heterogeneity in both mean and variance of ratings.

The dependent variable, the score of the jth joke in the ith sequence, y_{ij}, is represented by the following equation:

$$\begin{aligned}(3)\quad y_{ij} = {} & \beta_{0i} + \beta_{1i}y_{ij-1} + \beta_{2i}\text{FIRST}_i + \beta_{3i}\text{MAX}_{ij} + \beta_{4i}\text{MIN}_{ij} \\ & + \beta_{5i}\text{FIRSTNEG}_i + \beta_{6i}\text{FIRSTPOS}_i + \beta_{7i}\text{FIRSTNEG}_i \\ & \times \text{FIRST}_i + \beta_{8i}\text{FIRSTPOS}_i \times \text{FIRST}_i + \beta_{9i}\text{POSITION}_i \\ & + \beta_{10i}\text{POSITION}_i \times \text{FIRST}_i + \beta_{11i}\text{POSITION}_i \times \text{MAX}_i \\ & + \beta_{12i}\text{POSITION}_i \times \text{MIN}_i + \varepsilon_{ij}.\end{aligned}$$

The variables in the model are defined as follows. First, y_{ij-1} represents the score of the preceding joke in the sequence, or the j – 1th joke in the ith sequence. Just as with the previous model, a positive coefficient on this parameter reflects a positive correlation, indicating assimilation, and a negative coefficient indicates contrast. We expect the sign on β_{1i} (the parameter for score of the previous joke) to be negative, implying that the score of the focal joke contrasts with this value. We include the score of the first joke in the ith sequence directly as an independent variable (FIRST_i) to assess whether jokes in the sequence assimilate to the first joke, consistent with our previous results. The dummy variable FIRSTNEG_i equals 1 if the first stimulus is extremely negative, and 0 otherwise. Likewise, the dummy variable FIRSTPOS_i equals 1 if the first stimulus is extremely positive, and 0 otherwise. An interaction between one of these variables with the score of the first joke indicates that the first joke being an extreme has an exacerbating or attenuating effect on the observed effect of the first stimulus. The definitions of MAX_{ij} and MIN_{ij} are the same as in the beer model: the highest- and lowest-scoring joke, respectively, in the ith sequence leading up to the jth joke. POSITION is the ordinal position in which the joke occurs within the sequence. POSITION × FIRST, POSTION × MAX, and POSITION × MIN are the interaction terms between a joke's position and the first, highest-scoring, and lowest-scoring joke, respectively.

Individual heterogeneity. The vector of coefficients β_i represents sequence- (i.e., individual-) level random effects in which the intercept term β_{0i} indicates sequence-level intercepts, representing sequence-level assignment effects. The regression equations are related through a common prior distribution on the $\{\beta_i\}$. We take into account both observed participant-specific heterogeneity (i.e., the demographics age and gender and the individual variables expertise and mood) and unobserved participant-specific heterogeneity. We also include two dummy variables to capture differences, if any, between the three independent sources of participants in our data. To investigate how the participant-specific variables affect the magnitude of context effects, we specify the following multivariate regression:

$$\begin{aligned}(4)\quad \beta_{ki} = {} & \delta_0 + \delta_{k1}\text{AGE}_i + \delta_{k2}\text{GENDER}_i + \delta_{k3}\text{EXPERTISE}_i \\ & + \delta_{k4}\text{MOOD}_i + \delta_{k5}\text{SOURCE1}_i + \delta_{k6}\text{SOURCE2}_i + \nu_{ki}; \\ & \nu_{ki} \sim \text{i.i.d. } N(0, V_\beta).\end{aligned}$$

The k individual-specific parameters $\{\beta_{ki}\} = \{\beta_{0i}\ \beta_{1i}\ \beta_{2i}\ \beta_{3i}\ \beta_{4i}\ \beta_{5i}\ \beta_{6i}\ \beta_{7i}\ \beta_{8i}\ \beta_{9i}\ \beta_{10i}\ \beta_{11i}\ \beta_{12i}\}$ from Equation 3 are functions of the observable individual characteristics of expertise, age, gender, mood, and subject source. The coefficients δ_{k1}, δ_{k2}, δ_{k3}, δ_{k4}, δ_{k5}, and δ_{k6} indicate how a participant's age, gender, mood, expertise, and the source of participants modify the coefficients of the covariates in Equation 3. The variance–covariance matrix V_β determines the spread of the unobserved component of heterogeneity following notation from Rossi, Allenby, and McCulloch (2005) and as used in the first estimation. For the Gibbs sampler estimation, we use starting values obtained from an initial run of 100,000 draws, with the priors set to A = .5 × I and $\nu = 5 \times k + 3$, where k = 13 (the number of parameters in the model). We confirmed convergence using the Heidelberger and Welch convergence diagnostic as implemented

by the Bayesian output analysis program in R, and we used a chain of 200,000 draws for modeling the parameters' posterior distributions.

Results

We present the estimated posterior means and 99% confidence intervals in Table 3, Panels A and B. For ease of comprehension, we report the main effects in Panel A and the interaction effects (moderating effects of mood and expertise on the context effects) in Panel B. Although there is a negative main effect for age such that older people seemed to assign lower scores on average, the interaction effects for both demographic variables age and gender as well as those including the three sources of participants are not significant and are therefore not included in Table 3, Panel B.

Assimilation effects and extreme first experiences. The posterior mean of the first-joke coefficient is positive and significant ($\beta_{2i} = 3.27$, $p < .05$). The observed assimilation to the first experience is consistent with our beer data. The evaluation of the first joke, even in cases with extremely positive and negative first experiences, seems to act as an anchor and set the area of the scale used throughout the judging task. Moreover, the extremeness of the first experience plays a moderating role. The interaction of the first extreme positive dummy variable with the score of the first joke is negative ($\beta_{8i} = -10.48$, $p < .01$), indicating that extreme negatives also cause a contrast effect and can result in an overall attenuation of the first position assimilation effect.

Contrast effects. The coefficient on the score of the previous joke is significant and negative ($\beta_{1i} = -7.47$, $p < .01$), indicating contrast between adjacent stimuli. The coefficients on the score of the running maximum and minimum are also significant and negative ($\beta_{3i} = -2.41$, $p < .05$ and $\beta_{4i} = -3.25$, $p < .01$, respectively), indicating contrast with extremes and replicating the results obtained previously with the beer data. Importantly, we again find evidence of contrast effects and assimilation effects occurring simultaneously to different preceding evaluations within a sequence.

*Sequence position effect*s. The coefficient for the POSITION variable is not significant, implying that enjoyment of a joke seems to be independent of its particular position (e.g., third, fourth) in the sequence, and there are no observable satiation effects (Rolls et al. 1981; Rolls, Van Duijvenvoorde, and Rolls 1984). Position does affect the relative influence of the context effects. The later in the sequence that the focal joke appears, the greater the assimilation effect to the first score (POSITION × FIRST, $\beta_{10i} = 3.33$, $p < .05$). We also observe a reduction in contrast effects to the score of the maximum as the respondent moves down the sequence (POSITION × MAX, $\beta_{11i} = 10.8$, $p < .01$; positive). The pervasiveness of the first stimulus as a standard even as participants continue moving through the sequence speaks to the robustness of the assimilation effect. It also suggests that the first experience may be relied on to calibrate responses to subsequent experiences, and the repeated use of this calibration may make the assimilation stronger as respondents move down the sequence.

Expertise. We find a positive main effect for expertise, indicating that experts tend to assign higher scores on average ($\delta_{0\_3} = 2.85$, $p < .05$). Experts also show satiation effects down the sequence (POSITION – EXPERTISE, $\delta_{9\_3} = -3.02$, $p < .05$). More importantly, expertise significantly moderates the extent to which context effects influence sequential evaluations; experts experience *greater* context effects. Experts' evaluations assimilate more to the score of the first joke (FIRST – EXPERTISE, $\delta_{2\_3} = 3.91$, $p < .05$) and contrast more with both the score of the previous joke and the score of the running maximum (y_{ij-1} – EXPERTISE, $\delta_{1\_3} = -4.32$, $p < .01$; MAX – EXPERTISE, $\delta_{3\_3} = -2.79$, $p < .05$). It seems that experts are more susceptible than novices to making comparisons with previous evaluations.

This result is consistent with research that documents how experts often cannot ignore (and thus, overweight) additional information even when it is in their best interest to do so (Nam and Sternthal 2008). This is commonly referred to as the "curse of knowledge" effect (Camerer, Loewenstein, and Weber 1989; Hall, Ariss, and Todorov 2007). Compared with our joke novices, our joke experts seem to put more weight on prior evaluations, thereby increasing the likelihood of being influenced by context effects. Experts also are influenced by truly negative experiences. The contrast to the running minimum seems to be attenuated for experts (MIN – EXPERTISE, $\delta_{4\_3} = 7.05$, $p < .01$), and an extreme negative in the first position tends to have a negative impact on scoring only for experts (FIRSTNEG – EXPERTISE, $\delta_{5\_3} = -3.69$, $p < .05$). Both of these context effects result in lower scores for experts than for novices.

Mood. Mood as an intercept is not significant; a positive or negative mood by itself does not affect the scoring of jokes. We asked respondents to report their mood before evaluating the jokes, so respondents could not have misattributed their enjoyment of the jokes to their current mood (Schwarz and Clore 1983). However, contrast effects with extremes are enhanced when participants are in a more negative mood (MAX – MOOD, $\delta_{3\_4} = 4.22$, $p < .01$; MIN – MOOD, $\delta_{4\_4} = 4.76$, $p < .05$). This result is consistent with research by Raghunathan and Irwin (2001) on mood-based assimilation. Their results suggest positive mood-based assimilation; in other words, positive moods attenuate contrast effects. Conversely, negative moods (the absence of a positive mood) result in enhanced contrast effects. Finally, a more negative mood further increases contrast effects with extremes as judges work their way down the sequence (POSITION × MAX – MOOD, $\delta_{11\_4} = 4.15$, $p < .01$; POSITION × MIN – MOOD, $\delta_{12\_4} = 4.78$, $p < .01$).

Discussion

Whereas mood and expertise each had their own significant impact on context effects, demographics such as age and gender and the source of participants did not. We conducted additional analyses to ensure the robustness of our findings. As in the beer data, we estimated the model on the third and subsequent positions in the sequence because of constraints imposed by our definitions of running maximum and running minimum. We did a separate estimation of the model on position 2 and greater to include jokes in the second position because an extreme first stimulus could potentially affect the second experience to a greater extent. Doing so automatically dropped the variables running maximum and running minimum from the model. The estimates are comparable, and we again find assimilation to the first joke

Table 3
PARAMETER ESTIMATES FOR JOKES DATA

A: Level 1: Main Effects

Parameter	*Variable*	*Variable Description*	*99% Posterior Interval*	*Posterior Estimate (SD)*
β_{0i}	Intercept	Intercept	(–3.35, 4.38)	.64 (1.55)
β_{1i}	y_{ij-1}	Score of previous joke	(–12.1, –2.33)	–7.47 (1.73)**
β_{2i}	FIRST	Score of first joke in sequence	(.39, 7.11)	3.27 (1.61)*
β_{3i}	MAX	Score of running maximum	(–5.43, .90)	–2.41 (1.34)*
β_{4i}	MIN	Score of running minimum	(–6.98, –.03)	–3.25 (1.26)**
β_{5i}	FIRSTNEG	Dummy = 1 if first joke is extremely negative, and 0 otherwise	(–3.41, 5.43)	1.85 (1.63)
β_{6i}	FIRSTPOS	Dummy = 1 if first joke is extremely positive, and 0 otherwise	(–4.30, 4.73)	.70 (1.69)
β_{7i}	FIRSTNEG × FIRST	Interaction of FIRSTNEG dummy with score of first joke	(–3.55, 3.55)	.05 (1.51)
β_{8i}	FIRSTPOS × FIRST	Interaction of FIRSTPOS dummy with score of first joke	(–14.5, –3.03)	–10.48 (1.43)**
β_{9i}	POSITION	Position of joke in sequence	(–4.17, 2.82)	–.59 (1.51)
β_{10i}	POSITION × FIRST	Interaction of position and score of first joke	(–1.32, 7.52)	3.33 (1.60)*
β_{11i}	POSITION × MAX	Interaction of position and score of running maximum	(3.05, 15.80)	10.80 (1.69)**
β_{12i}	POSITION × MIN	Interaction of position and score of running minimum	(–3.44, 3.50)	.02 (1.41)

B: Level 2: Moderating Effects of Mood and Expertise

Parameter	*Variable*	*Variable Description*	*99% Confidence Interval*	*Posterior Estimate (SD)*
δ_{0_3}	Intercept – EXPERTISE	Main effect of expertise on scores	(–.72, 6.32)	2.85 (1.44)*
δ_{0_4}	Intercept – MOOD	Main effect of mood on scores	(–4.42, 1.98)	–1.35 (1.34)
δ_{1_3}	y_{ij-1} – EXPERTISE	Moderation effect of expertise on (contrast with) score of previous joke	(–7.79, –.73)	–4.32 (1.48)**
δ_{1_4}	y_{ij-1} – MOOD	Moderating effect of mood on (contrast with) score of previous joke	(–4.95, 2.32)	–1.33 (1.40)
δ_{2_3}	FIRST – EXPERTISE	Moderating effect of expertise on (assimilation to) score of first joke	(–.14, 7.85)	3.91 (1.54)*
δ_{2_4}	FIRST – MOOD	Moderating effect of mood on (assimilation to) score of first joke	(–3.10, 3.61)	.62 (1.34)
δ_{3_3}	MAX – EXPERTISE	Moderating effect of expertise on (contrast with) score of running max	(–6.03, .84)	–2.79 (1.48)*
δ_{3_4}	MAX – MOOD	Moderating effect of mood on (contrast with) score of running max	(.21, 8.27)	4.22 (1.66)**
δ_{4_3}	MIN – EXPERTISE	Moderating effect of expertise on (contrast with) score of running min	(.09, 10.2)	7.05 (1.48)**
δ_{4_4}	MIN – MOOD	Moderating effect of mood on (contrast with) score of running min	(–1.05, 8.39)	4.76 (1.46)*
δ_{5_3}	FIRSTNEG – EXPERTISE	Moderating effect of expertise on effect of extremely negative first joke	(–7.91, 1.69)	–3.69 (1.65)*
δ_{5_4}	FIRSTNEG – MOOD	Moderating effect of mood on effect of extremely negative first joke	(–6.50, 2.06)	–1.79 (1.68)
δ_{6_3}	FIRSTPOS – EXPERTISE	Moderating effect of expertise on effect of extremely positive first joke	(–1.83, 5.88)	2.29 (1.40)
δ_{6_4}	FIRSTPOS – MOOD	Moderating effect of mood on effect of extremely positive first joke	(1.30, 9.54)	5.44 (1.31)**
δ_{7_3}	FIRSTNEG × FIRST – EXPERTISE	Moderating effect of expertise on assimilation to extremely negative first joke	(–4.58, 3.36)	–.37 (1.57)
δ_{7_4}	FIRSTNEG × FIRST – MOOD	Moderating effect of mood on assimilation to extremely negative first joke	(–7.03, 1.71)	–2.86 (1.62)
δ_{8_3}	FIRSTPOS × FIRST – EXPERTISE	Moderating effect of expertise on assimilation to extremely positive first joke	(–2.66, 5.49)	.87 (1.53)
δ_{8_4}	FIRSTPOS × FIRST – MOOD	Moderating effect of mood on assimilation to extremely positive first joke	(.57, 10.2)	6.11 (1.48)**
δ_{9_3}	POSITION – EXPERTISE	Moderating effect of expertise on position in sequence	(–6.51, .06)	–3.02 (1.41)*
δ_{9_4}	POSITION – MOOD	Moderating effect of mood on position in sequence	(–2.29, 5.43)	1.79 (1.52)
δ_{10_3}	POSITION × FIRST – EXPERTISE	Moderating effect of expertise down the sequence on assimilation to first joke	(–11.7, –1.60)	–7.77 (1.38)**
δ_{10_4}	POSITION × FIRST – MOOD	Moderating effect of mood down the sequence on assimilation to first joke	(–4.02, 3.12)	–.69 (1.48)
δ_{11_3}	POSITION × MAX – EXPERTISE	Moderating effect of expertise down the sequence on effect of running max	(–3.08, 3.90)	.87 (1.34)
δ_{11_4}	POSITION × MAX – MOOD	Moderating effect of mood down the sequence on effect of running max	(1.17, 9.73)	4.15 (1.56)**
δ_{12_3}	POSITION × MIN – EXPERTISE	Moderating effect of expertise down the sequence on effect of running min	(–3.38, 3.66)	.97 (1.38)
δ_{12_4}	POSITION × MIN – MOOD	Moderating effect of mood down the sequence on effect of running min	(.90, 8.11)	4.78 (1.42)**

*$p < .05$.
**$p < .01$.

and contrast with the previous joke. Results from other estimations are also qualitatively similar (e.g., using the measures for objective expertise in the main model instead of subjective expertise; these results are available from the authors upon request).

The results from this experiment extend the findings from our analysis of the beer data to another context while demonstrating a similar pattern of results for a very different kind of hedonic evaluation. The experimental paradigm also enables us to identify specific factors that moderate the impact of multiple reference points on sequential evaluation, including the role of an extreme first stimulus as well as the moderating effects of individual-level factors such as expertise and mood. Table 4 summarizes our main findings across the two data sets.

GENERAL DISCUSSION

Theoretical Implications

In this research, we find simultaneously occurring assimilation and contrast effects to multiple reference points that systematically occur during the evaluation of a sequence of hedonic experiences. Although much of the research on context effects in marketing presupposes that only one or the other (assimilation or contrast) can occur, Mussweiler and Strack (2000, p. 35) propose that "the very same comparison can lead to both assimilation and contrast." However, they observe only one type of effect associated with a single self-evaluative judgment. For example, in one of their studies, when participants were asked to compare themselves with a high athletic standard, they assimilated to the standard if asked to describe their athleticism in words (an objective judgment) but contrasted away from the standard if asked to provide a numeric evaluation (subjective judgment scales). The authors argue that whether an assimilation or contrast effect is observed depends on the type of judgment elicited, which dictates whether the assimilative consequence (accessibility of standard-consistent knowledge) or the contrastive consequence (serving as a relevant reference point) dominates. In contrast, we observe assimilation and contrast effects simultaneously to different reference points, and we find assimilation effects even though we always elicit subjective judgments using numeric evaluation scales.

Table 4
SUMMARY OF KEY FINDINGS FOR SEQUENTIAL HEDONIC EVALUATION: STUDIES 1–2

Key Finding	*Study 1*	*Study 2*
Main Assimilation Effects		
Experiences assimilate to the first in the sequence	Yes	Yes
Main Contrast Effects		
Contrast with immediate predecessor	Yes	Yes
Contrast with extremes inside the sequence	Yes	Yes
Extreme first scores simultaneously cause contrast	Not tested	Yes
Position Effects		
Experiences in first position receive positive bump	Yes	Not tested
Context effects reduced by subcategory change	Yes	Not tested
Assimilation to first increases down sequence	Not found	Yes
Satiation causes decreasing scores down sequence	Not found	Found only for experts
Expertise		
Experts assign higher scores on an average	Not tested	Yes
Experts demonstrate greater context effects	Not tested	Yes
Mood		
Negative moods increase contrast effects	Not tested	Yes
Contrast to extremes (max and min) down sequence is further reduced if mood is more positive	Not tested	Yes

Note that our work differs from Mussweiler and Strack (2000) in many respects. First, the primary intent of the current research is to document how people assimilate to one reference point while simultaneously contrasting away from a different, distinct reference point. In contrast, Mussweiler and Strack claim that opposing comparative processes are simultaneously at play for a single reference point and observe a different comparative effect depending on the type of judgment elicited. Second, our focus is on sequences of hedonic experiences, whereas Mussweiler and Strack's focus is how people view themselves relative to someone else. Third, whereas Mussweiler and Strack explicitly instructed respondents to make comparisons, our effects seem to emerge spontaneously, without explicit instructions. This reassures us that our effects are not dependent on the procedural paradigm. Fourth, we document the same pattern of effects in both the lab (jokes) and the real world (beers). This finding is especially valuable given the importance of assessing external validity and the need for more work in the field that does so (Lynch 1999; Winer 1999).

We view our primary contribution as having identified and documented these effects. Given the limitations of our data, we remain agnostic with respect to a specific underlying mechanism or process. We believe it is informative to raise possible explanations when appropriate, as we have done, but we acknowledge the speculative nature of our assertions, especially in light of the historic difficulty associated with pinpointing with any certainty the mechanism(s) responsible for a specific comparative effect (Mussweiler 2003). Still, our results should be compelling to scholars interested in the ongoing debate regarding which process(es) (e.g., numeric priming, anchoring and adjustment, selective accessibility, scale distortion) can account for the many types of comparative effects scholars have observed (Mochon and Frederick 2013).We also recognize the inability to resolve whether some or all of our effects are the result of changes in judges' mental representations of the evaluation versus changes in how judges used the response scale to report those evaluations.

We believe that prior research in marketing has neglected the possibility of simultaneously occurring comparative effects to multiple references points for two reasons—one being theoretical and the other, methodological. Theoretically, scholars in marketing have tended to treat assimilation and contrast as an "either/or" phenomenon. This is likely due to the researchers' goals, which have focused primarily on identifying which effect occurs in situations when the target could not exist in two disparate "states" simultaneously. For example, when exploring how the role of category member-

ship affects comparative evaluations (assimilation/contrast), a target is not typically viewed as existing both within and outside a category. Or, if testing the effect of similarity, it is difficult to view a target stimulus as being perceived to be both similar and dissimilar to the same reference standard in the same context. Unlike previous research, we were not constrained by this type of specific hypothesis testing while setting out to document the effects we observe.

Methodologically, much of the previous research has relied on experimental data in which an initial stimulus is created to form the context and a subsequent stimulus is created to form the target of the evaluation. For example, in Zellner et al. (2003), one group tasted a full-strength fruit juice followed by the same juice diluted with water. In Zellner et al. (2010), respondents viewed and rated five paintings from Francisco Goya's Dark Period (hedonically negative context stimuli) before viewing and rating two pastoral paintings by Goya (hedonically neutral test stimuli). This type of operationalization is limiting in that it does not allow the researcher to test for both assimilation and contrast effects concurrently. In addition, these experiments test whether people will experience assimilation *or* contrast depending on the specific conditions set up in the study, whereas our studies test whether people experience *both* kinds of effects within unpreordained sequences of evaluations. Accordingly, our research offers both internal and external validity to the claim that both assimilation and contrast can occur to multiple reference points for individual targets evaluated within a sequence of hedonic experiences. It is our hope that future studies might use a similar methodology to ensure testing for any and all comparative effects.

Managerial Implications

Evaluations of contesting stimuli often seem distinct and concrete, but often the scores are separated by very small margins in which underlying context effects may decide the winners and losers. Consider that in 33% of the flight contests at the Bluebonnet Brew-Off, one of the largest and most prestigious home brew competitions in the United States, first-place beers won by one point or less, and a difference of up to two points decided 58% of the winners. Given the substantial effects we find in our model, these statistics reveal that a sizeable proportion of the outcomes may have been decided by assimilation and contrast effects. Unbeknownst to many contest judges, wine enthusiasts, or even those of us flipping through television channels searching for a show to watch, our experiences and the order in which we experience them are likely to affect our evaluations and, therefore, our subsequent choices (Wirtz et al. 2003).

Our results contribute to an increasing literature stream documenting the ramifications inherent in how one presents a choice, referred to as the "choice architecture" (Thaler and Sunstein 2008). In our case, it involves constructing sequences for trial. Managers may have many potential goals when designing sequences of experiences. They may be trying to elicit the most accurate evaluations, enable the most fair evaluation of each item, maximize (or minimize) the overall enjoyment of the total experience, or even make one choice look more promising than the others. Our research provides insights for each of these objectives.

Accuracy. One goal may be to truly minimize context effects to arrive at an evaluation that most accurately represents an item's "real" score, if one exists. It is often in marketers' best interest for their customers to make the most accurate judgments, especially when there are long-term relationships between the marketer and the customer. For example, in both business-to-business and business-to-customer relationships in which customer satisfaction, trust, and repeat purchases are important, marketers want customers to provide accurate evaluations that facilitate the future delivery of better-customized products and services. In these situations, marketers should attempt to minimize context effects.

When addressing the issue of context effects, conventional wisdom dictates that more objective measures of comparative quality are secured when the sequence of stimuli are ordered randomly. Thus, randomization is preferred when multiple products are being assessed at one time. Our results show that for accuracy, randomization is not the panacea it seems to be. In this research, we explicitly model biases that occur as a result of sequential evaluation, irrespective of the nature of the sequence. In other words, these biases occur because subsequent items in a sequence always have preceding items that can influence them; context effects should therefore occur in every potential ordering of a sequence. Random ordering alone cannot eliminate such biases, although it gives each item an equal chance a priori of being hindered or helped.

An obvious recommendation to help increase accuracy with random sequences is to encourage repeat trials in new and differing orders. In the Bluebonnet Brew-Off, each judge might be asked to sample a flight multiple times in different sequences, or, more cost effectively, the two or three judges could each sample the same flight, but in different orders. Although we suspect that a greater sampling both in number and duration would help judges form more objective evaluations, when sequences and sample sizes are large, this method may become unwieldy. For example, it can take an hour for a judge to evaluate a single flight of 12 beers. Our research can suggest alternatives to this approach.

A promising avenue of further research would be to have people switch mental categories more often. If participants can focus on how judged items are fundamentally different and, thus, should not be compared, the prior items may become less contextually relevant. Typically, one might assume that accenting differences would lead to greater differences in scores and, thus, more contrast effects (Sherif and Hovland 1961). Yet the beer evaluation findings demonstrate that context effects are attenuated by subcategory change. If differences are used as cues that items must be judged individually (because they are in different categories), contrast effects should be attenuated. However, if judges focus on making comparison judgments drawing on these differences, contrast effects are likely to be enhanced. Perhaps something as simple as extending the time between trials could prompt judges to evaluate stimuli individually. Timing of trials and context effects is a possible solution that researchers can investigate in the future.

Another potential solution is to strive for a more relativistic accuracy. In this method, each item is evaluated directly against a category "exemplar," and the experience of the exemplar acts as a base-level experience (outside standard).

Relative comparisons may work if a goal is to compare only within a singular sequence or even across multiple sequences with similar characteristics. Consistently using the same exemplar as an immediate precursor, which is the strongest point of contrast, would be unlikely to eradicate contrast effects, but it could perhaps convert the contrast effect from an unknown, varying magnitude and direction into a more systematic bias. Evaluations would become more accurate and stable, at least relative to one another. For example, a person tasting merlot wines could use a Kendall-Jackson merlot—a "good" wine with a *Wine Enthusiast* score of 86—as the "baseline" wine that is sipped to recalibrate a baseline between tastes. It is important to note that a neutral cracker, which should be consumed before each tasting to cleanse the palate, does not perform the same task because it is not judged and therefore does not attenuate the context effects (at least in our beer study).

Fairness. It is important to disentangle fairness from accuracy. Accuracy aims at measuring a true, objective score or judgment. Fairness is easier to obtain because it aims at an equality of measurement within a choice set; all evaluations can be biased as long as each evaluation is biased equally or with equal likelihood. Although random ordering does not diminish the significant effect of context on determining the outcome of a competition, it may make the judging in the competition more "fair." When marketers are involved in competitions, especially those that may be considered to have subjective judging (e.g., reality show contests, advertisement entries for the Doritos "Crash the Super Bowl" contest), fairness is often as important as, or more so than, accuracy. A public display of order randomization can increase the belief that each entry received a "fair shot" in these situations.

Enjoyment. As diners and viewers of reality competition programs, we may strive to make accurate decisions regarding our choices. Chefs and television executives, however, likely have another goal in mind. They are trying to get their audiences to maximize their overall evaluation or enjoyment of the sequence of experiences. Retrospective evaluations have been shown to be predicted fairly well by some average of the peak rating and the final rating (i.e., the "peak" and the "end") of an extended temporal experience (Kahneman et al. 1993; Kubovy 1999). Peak–end theory predicts how retrospective, not real-time, evaluations are formed. However, this retrospective evaluation is composed of a sequence of real-time evaluations that may be influenced by the multiple comparative effects we identify. Systematically steering the effects to enhance positively valenced peaks and ends should lead to greater overall evaluations.

In addition, an extremely high score in the first experience is important. A chef may want evaluations of the first course to be as high as possible because all others in the sequence will assimilate to this. Our beer research suggests that contestants in the first position often experience a slight positive lift. By putting one of the strongest items first, managers can ensure a strong positive assimilation effect. Our joke study further demonstrates that among experts, subsequent evaluations assimilate more strongly to extremely bad initial experiences, which means that starting off with a negative experience is potentially doubly harmful. In addition to the importance of beginning with a strong positive experience, we recommend that other items that are predicted to be experienced highly be preceded by the lowest expected items. This should lead to the greatest positive contrast effects and, thus, the highest sequential peaks.

Alternatively, there could be occasions when negative contrasts in the sequence are unwanted, possibly for issues of fairness. This situation would promote the desire to present options in increasing sequences. Further research should explore whether alternating high/low choices or presenting an increasing sequence leads to stronger contrast effects, especially when incorporating our findings with research on moment-to-moment evaluations, duration neglect, and the peak–end rule. After all, peak–end theory addresses the holistic judgments of entire experiences, whereas our research pertains to sequences of individual judgments. Still, inasmuch as summary judgments are composed of individual events, our work may have a strong influence on this avenue of inquiry.

Influencing evaluations. In many purchase situations, managers may not care if customers like all the products as long as they like at least one of the offerings enough to purchase it. Managers may not want to strive for fairness or accuracy but instead intend to "nudge" people toward a certain choice (Thaler and Sunstein 2008). In such situations, it may be best to heighten contrast with other offerings within the choice set. For example, a salesperson at a perfume counter may want to include a wide range of perfumes for a customer to sample, including some the salesperson expects would not be preferred. By including these "bad" perfumes in the choice set, the salesperson provides points of contrast that should increase the evaluations of the "good" perfumes, thereby increasing the likelihood of a purchase. We caution the salesperson to begin with one of the better perfumes because the first position usually experiences a slight lift in evaluation, and this would provide a positive assimilation effect for the entire sequence. Stressing the similarity of the choices in the category may also increase the likelihood for contrast effects.

The ways that marketers can construct sequences of experiences are not limited to menu items, movie trailers, or performance orders. Even the construction of retail spaces can allow for the careful control of sequential experience. For example, both IKEA and Central Market, a gourmet food chain in Texas, design their stores so that customers must walk through the store using a prescribed path from department to department, effectively turning the shopping trip into a sequence of shopping experiences in which context effects could occur. The possibilities for marketers are limitless, and we hope that future empirical studies will further define and document the pervasiveness of these effects. We view our research as a critically important first step.

APPENDIX: EXTREMELY POSITIVE AND NEGATIVE JOKE STIMULI

Extremely Positive Joke

Three women die together in an accident and go to heaven. When they get there, St. Peter says, "We only have one rule here in heaven: don't step on the ducks!" So they enter heaven, and sure enough, there are ducks all over the place. It is almost impossible not to step on a duck, and although they try their best to avoid them, the first woman

accidentally steps on one. Along comes St. Peter with the ugliest man she ever saw. St. Peter chains them together and says, "Your punishment for stepping on a duck is to spend eternity chained to this ugly man!" The next day, the second woman accidentally steps on a duck, and along comes St. Peter, who doesn't miss a thing. With him is another extremely ugly man. He chains them together with the same admonishment as for the first woman. The third woman has observed all this and, not wanting to be chained for all eternity to an ugly man, is very, *very* careful where she steps. She manages to go months without stepping on any ducks, but one day St. Peter comes up to her with the most handsome man she has ever laid eyes on—very tall, long eyelashes, muscular, and thin. St. Peter chains them together without saying a word. The happy woman says, "What did I do to deserve being chained to you for all of eternity?" The guy says, "I don't know about you, but I stepped on a duck."

Extremely Negative Joke

Q: How do you smuggle an elephant across the border?
A: Put a slice of bread on each side, and call him "lunch."

REFERENCES

Allenby, Greg M. and Peter E. Rossi (1999), "Marketing Models of Consumer Heterogeneity," *Journal of Econometrics*, 89 (1/2), 57–78.

Ariely, Dan, George Loewenstein, and Drazen Prelec (2003), "'Coherent Arbitrariness': Stable Demand Curves Without Stable Preferences," *Quarterly Journal of Economics*, 118 (1), 73–106.

Bjork, Staffan and Jussi Holopainen (2004), *Patterns in Game Design*. Hingham, MA: Charles River Media Inc.

Brinberg, David L. and Joseph Edward McGrath (1985), *Validity and the Research Process*. Beverly Hills, CA: Sage Publications.

Bruine de Bruin, Wandi and Gideon Keren (2003), "Order Effects in Sequentially Judged Options Due to the Direction of Comparison," *Organizational Behavior and Human Decision Processes*, 92 (1/2), 91–101.

Camerer, Colin, George Loewenstein, and Martin Weber (1989), "The Curse of Knowledge in Economic Settings: An Experimental Analysis," *Journal of Political Economy*, 97 (5), 1232–54.

Chapman, Gretchen B. and Eric J. Johnson (1994), "The Limits of Anchoring," *Journal of Behavioral Decision Making*, 7 (4), 223–42.

—— and —— (1999), "Anchoring, Activation, and the Construction of Values," *Organizational Behavior and Human Decision Processes*, 79 (2), 115–53.

Damisch, Lysann, Thomas Mussweiler, and Henning Plessner (2006), "Olympic Medals as Fruits of Comparison? Assimilation and Contrast in Sequential Performance Judgments," *Journal of Experimental Psychology: Applied*, 12 (3), 166–78.

Dato-on, Mary C. and Robert F. Dahlstrom (2003), "A Meta-Analytic Investigation of Contrast Effects in Decision Marketing," *Psychology and Marketing*, 20 (8), 1–21.

Davis, Donna F., Susan L. Golicic, Courtney N. Boerstler, Sunny Choi, and Hanmo Oh (2013), "Does Marketing Research Suffer from Methods Myopia?" *Journal of Business Research*, 66 (9), 1245–50.

Deighton, John, Debbie MacInnis, Ann McGill, and Baba Shiv (2010), "Broadening the Scope of Consumer Research," *Journal of Consumer Research*, 36 (1), 5–7.

Ehrenberg, A.S.C. (1995), "Empirical Generalisations, Theory, and Method," *Marketing Science*, 14(3), G20–G28.

Gelfand, Alan E. and Adrian F.M. Smith (1990), "Sampling-Based Approaches to Calculating Marginal Densities," *Journal of the American Statistical Association*, 85 (410), 398–409.

Hall, Crystal C., Lynn Ariss, and Alexander Todorov (2007), "The Illusion of Knowledge: When More Information Reduces Accuracy and Increases Confidence," *Organizational Behavior and Human Decision Processes*, 103 (2), 277–90.

Herr, Paul M. (1986), "Consequences of Priming: Judgment and Behavior," *Journal of Personality and Social Psychology*, 51 (6), 1106–1115.

——, Steven J. Sherman, and Russell H. Fazio (1983), "On the Consequences of Priming: Assimilation and Contrast Effects," *Journal of Experimental Social Psychology*, 19 (4), 323–40.

Hubbard, Raymond and R. Murray Lindsay (2002), "How the Emphasis on 'Original' Empirical Marketing Research Impedes Knowledge Development," *Marketing Theory*, 2 (4), 381–402.

Kahneman, Daniel, Barbara L. Fredrickson, Charles A. Schreiber, and Donald A. Redelmeier (1993), "When More Pain Is Preferred to Less: Adding a Better End," *Psychological Science*, 4 (6), 401–405

Kardes, Frank R. and Paul M. Herr (1990), "Order Effects in Consumer Judgment, Choice, and Memory: The Role of Initial Processing Goals," in *Advances in Consumer Research*, Vol. 17, Marvin E. Goldberg, Gerald Gorn, and Richard W. Pollay, eds. Provo, UT: Association for Consumer Research, 541–46.

Kleiser, Susan B. and Susan P. Mantel (1994), "Consumer Expertise: A Scale Development," in *Enhancing Knowledge Development in Marketing*, Vol. 5, Andrew Mitchell and Ravi Achrol, eds. Chicago: American Marketing Association, 20–25.

Koop, Gregory. J. and Joseph G. Johnson (2012), "The Use of Multiple Reference Points in Risky Decision Making," *Journal of Behavioral Decision Making*, 25 (1), 49–62.

Kubovy, Michael (1999), "On the Pleasures of the Mind," in *Well-Being: The Foundations of Hedonic Psychology*, Daniel Kahneman, Ed Diener, and Norbert Schwarz, eds. New York: Russell Sage Foundation.

Lehmann, Donald R., Leigh McAlister, and Richard Staelin (2011), "Sophistication in Research in Marketing," *Journal of Marketing*, 75 (July), 155–65.

Lynch, John G., Jr. (1999), "Theory and External Validity," *Journal of the Academy of Marketing Science*, 27 (3), 367–76.

—— (2011), "Substantive Consumer Research," in *Advances in Consumer Research*, Vol. 38, Darren W. Dahl, Gita V. Johar, and Stijn M.J. van Osselaer, eds. Duluth, MN: Association for Consumer Research.

——, Joseph W. Alba, Aradhna Krishna, Vicki G. Morwitz, and Zeynep Gürhan-Canli (2012), "Knowledge Creation in Consumer Research: Multiple Routes, Multiple Criteria," *Journal of Consumer Psychology*, 22 (July) 473–85.

——, Dipankar Chakravarti, and Anusree Mitra (1991), "Contrast Effects in Consumer Judgments: Changes in Mental Representations or in the Anchoring of Rating Scales," *Journal of Consumer Research*, 18 (3), 284–97.

MacFie, Halliday J., Nicholas Bratchell, Keith Greehoff, and Lloyd V. Vallis (1989), "Design to Balance the Effect of Order of Presentation and First-Order Carryover Effect in Hall Tests," *Journal of Sensory Studies*, 4 (2), 129–48.

March, James G. and Zur B. Shapira (1992), "Variable Risk Preferences and the Focus of Attention," *Psychological Review*, 99 (1), 172–83.

Mayhew, Glenn E. and Russell S. Winer (1992), "An Empirical Analysis of Internal and External Reference Prices Using Scanner Data," *Journal of Consumer Research*, 19 (1), 62–70.

Mochon, Daniel and Shane Frederick (2013), "Anchoring in Sequential Judgments," *Organizational Behavior and Human Decision Processes*, 122 (1), 69–79.

Mussweiler, Thomas (2003), "Comparison Processes in Social Judgment: Mechanisms and Consequences," *Psychological Review*, 110 (3), 472–89.

——— and Fritz Strack (2000), "The 'Relative Self': Informational and Judgmental Consequences of Comparative Self-Evaluation," *Journal of Personality and Social Psychology*, 79 (1), 23–38.

Nam, Myungwoo and Brian Sternthal (2008), "The Effects of a Different Category Context on Target Brand Evaluations," *Journal of Consumer Research*, 35 (4), 668–79.

Novemsky, Nathan and Rebecca K. Ratner (2003), "The Time Course and Impact of Consumers' Erroneous Beliefs About Hedonic Contrast Effects," *Journal of Consumer Research*, 29 (4), 507–516.

Ordóñez, Lisa D., Terry Connolly, and Richard Coughlin (2000), "Multiple Reference Points in Satisfaction and Fairness Assessment," *Journal of Behavioral Decision Making*, 13 (3), 329–44.

Page, Benjamin I. and Robert Y. Shapiro (2010), *The Rational Public: Fifty Years of Trends in Americans' Policy Preferences*. Chicago: University of Chicago Press.

Page, Lionel and Katie Page (2010), "Last Shall Be First: A Field Study of Biases in Sequential Performance Evaluations on the Idol Series," *Journal of Economic Behavior & Organization*, 73 (2), 186–98.

Parducci, Alan (1964), "Category Judgment: A Range-Frequency Model," *Psychological Review*, 72 (6), 407–418.

Podsakoff, Philip M., Scott B. MacKenzie, Jeong-Yeon Lee, and Nathan P. Podsakoff (2003), "Common Method Biases in Behavioral Research: A Critical Review of the Literature and Recommended Remedies," *Journal of Applied Psychology*, 88 (5) 879–903.

Raghunathan, Rajagopal and Julie R. Irwin (2001), "Walking the Hedonic Product Treadmill: Default Contrast and Mood-Based Assimilation Effects in Judgments of Predicted Happiness with Target Product," *Journal of Consumer Research*, 28 (3), 355–68.

Redden, Joseph P. (2008), "Reducing Satiation: The Role of Categorization Level," *Journal of Consumer Research*, 34 (2), 624–34.

Reiss, Peter C. (2011), "Descriptive, Structural, and Experimental Methods in Marketing Research," *Marketing Science*, 30 (6), 950–64.

Rolls, Barbara J., Edmund T. Rolls, Edward A. Rowe, and Kevin Sweeney (1981), "Sensory Specific Satiety in Man," *Physiology and Behavior*, 27 (1), 137–42.

———, P.M. Van Duijvenvoorde, and Edmund T. Rolls (1984), "Pleasantness Changes and Food Intake in a Varied Four-Course Meal," *Appetite*, 5 (4), 337–48.

Rossi, Peter E., Greg M. Allenby, and Robert McCulloch (2005), *Bayesian Statistics and Marketing*. London: John Wiley & Sons.

Rota, Lauren M. and Debra A. Zellner (2007), "The Categorization Effect in Hedonic Contrast: Experts Differ from Novices," *Psychonomic Bulletin & Review*, 14 (1), 179–83.

Schwarz, Norbert and Herbert Bless (1992), "Constructing Reality and Its Alternatives: An Inclusion/Exclusion Model of Assimilation and Contrast Effects in Social Judgment," in *The Construction of Social Judgments*, Leonard L. Martin and Abraham Tesser, eds. Hillsdale, NJ: Lawrence Erlbaum Associates, 217–48.

——— and Gerald L. Clore (1983), "Mood, Misattribution, and Judgments of Well-Being," *Journal of Personality and Social Psychology*, 45 (3), 513–23.

——— and ——— (1996), "Feelings and Phenomenal Experiences," in *Social Psychology: Handbook of Basic Principles*, E. Tory Higgins and Arie Kruglanski, eds. New York: Guilford Press, 433–65.

Sherif, Muzafer and Carl I. Hovland (1961), *Social Judgment: Assimilation and Contrast Effects in Communication and Attitude Change*. New Haven, CT: Yale University Press.

Strong, Gordon and Steve Piatz (2008), "BJCP Exam Study Guide," (September).

Thaler, Richard H. and Cass R. Sunstein (2008), *Nudge: Improving Decisions About Health, Wealth, and Happiness*. New Haven, CT: Yale University Press.

Tversky, Amos and Daniel Kahneman (1974), "Judgment Under Uncertainty: Heuristics and Biases," *Science*, 185 (4157), 1124–31.

Wang, Xiao-Tian (2008), "Risk Communication and Risky Choice in Context: Ambiguity and Ambivalence Hypothesis," *Annals of the New York Academy of Sciences*, 1128 (April), 78–89.

Welch, Joe L. and Cathy O. Swift (1992), "Question Order Effects in Taste Testing of Beverages," *Journal of the Academy of Marketing Science*, 20 (3), 265–68.

Winer, Russell S. (1986), "A Reference Price Model of Brand Choice for Frequently Purchased Products," *Journal of Consumer Research*, 13 (2), 250–56.

——— (1999), "Experimentation in the 21st Century: The Importance of External Validity," *Journal of the Academy of Marketing Science*, 27 (3), 349–58.

Wirtz, Derrick, Justin Kruger, Christie Napa Scollon, and Ed Diener (2003), "What to Do on Spring Break? The Role of Predicted, On-Line, and Remembered Experience in Future Choice," *Psychological Science*, 14 (5), 520–24.

Wright, George, Gene Rowe, Fergus Bolger, and John Gammack (1994), "Coherence, Calibration, and Expertise in Judgmental Probability Forecasting," *Organizational and Human Decisions Processing*, 57 (1), 1–25.

Wyer, Robert S., Jr., and Thomas K. Srull (1989), *Memory and Cognition in Its Social Context*. New York: Lawrence Erlbaum Associates.

Zellner, Debra A., Ke'nesha Jones, Jennifer Morino, Elizabeth S. Cogan, and Emily M. Jennings (2010), "Increased Hedonic Differences Despite Increases in Hedonic Range," *Attention, Perception & Psychophysics*, 72 (5) 1261–65.

———, Brett B. Kern, and Scott Parker (2002), "Protection for the Good: Subcategorization Reduces Hedonic Contrast," *Appetite*, 38 (3), 175–80.

———, Elizabeth Rohm, Terri Bassetti, and Scott Parker (2003), "Compared to What? Effects of Categorization on Hedonic Contrast," *Psychonomic Bulletin and Review*, 10 (2), 468–73.