Volume 23, Issue 4, October 2013

JCP

JOURNAL OF CONSUMER PSYCHOLOGY

The Official Journal of The Society for Consumer Psychology

EDITORIAL

RESEARCH ARTICLES

RESEARCH REPORTS

RESEARCH DIALOGUE

ELSEVIER
ISSN 1057-7408

Available online at www.sciencedirect.com

ScienceDirect

Journal of Consumer Psychology 23, 4 (2013) 464 – 482

JCP JOURNAL OF CONSUMER PSYCHOLOGY

Research Article

Word-of-mouth and the forecasting of consumption enjoyment

Stephen X. He [a], Samuel D. Bond [b,*]

[a] Manhattan College, 4513 Manhattan College Parkway, Riverdale, NY 10471, USA
[b] Georgia Institute of Technology, 800 West Peachtree St NW, Atlanta, GA 30308-0520, USA

8 November 2011; 6 April 2013; 9 April 2013
Available online 15 April 2013

Abstract

The digital era has permitted rapid transfer of peer knowledge regarding products and services. In the present research, we explore the value of specific types of word-of-mouth information (numeric ratings and text commentary) for improving forecasts of consumption enjoyment. We present an anchoring-and-adjustment model in which the relative forecasting error associated with ratings and commentary depends on the extent to which consumer and reviewer have similar product-level preferences. To test our model, we present four experiments using a range of hedonic stimuli. Implications for the provision of consumer WOM are discussed.

Keywords: Word-of-mouth; Affective forecasting; Similarity; Preference heterogeneity

> "Enjoying the joys of others and suffering with them — these are the best guides for man."
>
> [Albert Einstein]

Introduction

For many consumer choices, successful decision making depends on the ability to accurately predict future consumption experience. Unfortunately, an abundance of evidence has revealed that individuals are generally poor at estimating their future affective states (e.g., Kahneman & Snell, 1992; Wilson & Gilbert, 2003). In principle, modern communication environments offer a means of facilitating the consumer forecasting process, by increasing access to word-of-mouth (WOM) through which product-relevant information is transmitted between consumers (Brown & Reingen, 1987). However, despite its prevalence and assumed benefits, there is scant empirical evidence that WOM actually enables consumers to make better forecasts. Moreover, there is little understanding of conditions under which different forms of WOM are more useful for forecasting purposes. The present research addresses these issues.

Among the myriad varieties of product-relevant WOM, we focus on that subset in which consumers present their own, usage-based experience and opinions directly. From the perspective of a prospective consumer, such WOM represents a form of 'surrogate' information, provided by a peer consumer who has experienced the product first-hand (Gilbert, Killingsworth, Eyre, & Wilson, 2009; Solomon, 1986). However, the information itself may vary widely, from a simple summary evaluation ("I hated the movie!") to underlying descriptive or explanatory commentary ("The plot was OK, but the acting was atrocious!..."), to some combination of the two. Our research question concerns the conditions under which each type of information (or their combination) will be beneficial to prospective consumers, by helping them to forecast their own product enjoyment.

To address this question, we focus on consumer reviews of the type found at online retailers or third-party platforms, which can be decomposed into two constituent elements: summary evaluations (i.e., ratings) and review commentary (i.e., text reviews). A number of scholarly investigations have documented the influence of product ratings on sales (Chevalier & Mayzlin, 2006; Liu, 2006; Moe & Trusov, 2011), and a separate literature has investigated the economic impact of commentary (Archak,

* Corresponding author. Fax: +1 404 894 6030.
E-mail addresses: stephen.he@mgt.gatech.edu (S.X. He), sam.bond@mgt.gatech.edu (S.D. Bond).

1057-7408/$ -see front matter © 2013 Society for Consumer Psychology. Published by Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jcps.2013.04.001

Ghose, & Ipeirotis, 2011; Park, Lee, & Han, 2007), but there has been almost no research directly comparing these types of information on consumer outcomes. In contrast, we explicitly adopt a consumer perspective. Extending recent work on the subjective 'helpfulness' of consumer review content (Mudambi & Schuff, 2010; Sen & Lerman, 2007), we focus directly on the utilization of WOM to predict future enjoyment and satisfaction.

Although numeric ratings and commentary both provide useful information about the experience of peer consumers, their relative value is unclear. Intuitively, marketers and consumers might expect a rating to be less useful than a commentary (Archak et al., 2011), as the latter provides both objective and subjective information, allowing prospective consumers to simulate their product experience in advance (Adaval & Wyer, 1998). However, research in affective forecasting reveals a variety of biases and limitations which cast doubt on this assumption (Wilson & Gilbert, 2003; Wood & Bettman, 2007). Moreover, although it may be assumed that forecasts will be most accurate when a reviewer's rating and commentary are presented together (as is the case on most real-world platforms), consumer researchers have long challenged the notion that "more information is better" (Jacoby, Speller, & Kohn, 1974; Keller & Staelin, 1987). It therefore remains an open question whether ratings, commentary, or their combination will produce the most accurate forecasts.

In the sections that follow, we address a previously unexplored area within consumer affective forecasting, by examining how consumers utilize word-of-mouth to predict their product enjoyment. To do so, we present an anchoring-and-adjustment framework in which a critical factor is the extent to which consumer and reviewer share similar product-level preferences. This framework allows us to examine the relative value of ratings, commentary, or their combination for making affective forecasts. To support our framework, we present four experimental studies which utilize different product categories and vary preference similarity both directly and indirectly. We show that the forecasting value of ratings declines substantially when consumers encounter reviewers having dissimilar preferences, whereas the value of commentary is largely unaffected by preference similarity. Moreover, a combination of rating and commentary together is sometimes less useful than either alone. We conclude by offering implications for the use of WOM to improve real-world consumer decision outcomes.

Conceptual background

Word-of-mouth as forecasting aid

The ability of consumers to accurately forecast their future consumption experience has notable psychological and economic consequences. Overestimation of future enjoyment may result in post-purchase regret and dissatisfaction, while underestimation may result in forgone opportunities for both consumer and marketer. Therefore, both parties stand to gain from the alignment of forecast with actual experience, and the topic has received substantial scholarly attention (Hoch, 1988; Loewenstein & Adler, 1995; Patrick, MacInnis, & Park, 2007; Wang, Novemsky, & Dhar, 2009). A robust finding of this work is that individuals are poor at making affective forecasts, particularly for hedonic events (Billeter, Kalra, & Loewenstein, 2011; Kahneman & Snell, 1992; Read & Loewenstein, 1995; Simonson, 1990; Wilson, Wheatley, Meyers, Gilbert, & Axsom, 2000; Wood & Bettman, 2007). Forecasting errors are most commonly attributed to faulty simulation of future experience (Gilbert & Wilson, 2007; Zhao, Hoeffler, & Dahl, 2009), and prescriptive advice often aims at improving the simulation process.

In keeping with broader research on the use of peer knowledge for personal prediction (Gershoff, Mukherjee, & Mukhopadhyay, 2003; Gilbert et al., 2009), our work highlights the role of WOM as a means of improving consumers' ability to forecast their enjoyment of goods and services in the marketplace. We focus in particular on online WOM, which has gained increasing attention in consumer research. A great deal of interest has been directed towards the various drivers of online WOM (Berger & Schwartz, 2011; De Angelis et al., 2012), its diverse effects on decision processing (Chan & Cui, 2011; Weiss, Lurie, & MacInnis, 2008; Zhao & Xie, 2011) and consequences for purchase behavior (Chevalier & Mayzlin, 2006; Zhu & Zhang, 2010). Surprisingly, although recent work has addressed the subjective value of WOM in terms of perceived 'helpfulness' (Mudambi & Schuff, 2010; Schindler & Bickart, 2012; Sen & Lerman, 2007), almost no attention has been paid to its more direct value in improving consumer decision outcomes.

Modern consumer WOM takes place over an evolving variety of channels that vary in scale, scope, and efficiency (blogs, social networks, mobile platforms, etc.), and the content of WOM may be categorized in numerous ways (informative vs. persuasive, first-hand vs. second-hand, positive vs. negative, etc.). For present purposes, we restrict our focus to instances in which WOM is utilized by consumers to share their own usage experience and opinions directly with their audience, e.g., consumer reviews of the type commonly available at online retailers and third-party review forums; however, the logic developed below can be extended to other channels (and we return to this issue later). Reviews are especially suited to our inquiry because they contain two distinct components, each of which has been widely studied (Chevalier & Mayzlin, 2006; Dellarocas, Zhang, & Awad, 2007; cf. Park et al., 2007). First, review platforms typically request that reviewers provide an overall product evaluation in the form of a numeric rating, often expressed symbolically ('stars,' etc.). Although consumers may disagree on the perceptual meaning of specific ratings, they do generally know the range of possible values and recognize that larger values connote more positive evaluations. Under ideal conditions, therefore, an overall rating conveys the reviewer's opinion accurately, with minimal effort required from the reader. Second, platforms often allow reviewers to provide text commentary that describes their experience with the product and explains their subsequent evaluation. In contrast to an overall rating, a commentary provides a richer context, often including vivid and concrete content that allows readers to mentally simulate their own potential product experience (Adaval & Wyer, 1998; Dickson, 1982). Although the helpfulness of a commentary varies by depth and readability (Archak et al., 2011; Mudambi & Schuff, 2010), it typically contains both objective and subjective content relevant to the decision. Moreover, a commentary often provides reasons underlying the author's evaluation, which may in turn be utilized by the reader to resolve decision conflict (Shafir, Simonson, & Tversky, 1993).

Given these differences, it may be natural for consumers to assume that a commentary is more helpful than a simple overall rating for prediction.[1] However, research on the communication of experiences casts doubt on the validity of this assumption. As a written explanation of a reviewer's experience, a commentary is likely to overemphasize certain aspects that are easier to recall or verbalize (Schooler & Engstler-Schooler, 1990), and may also contain reasoning that is ad hoc or inconsistent with the reviewer's attitude (Sengupta & Fitzsimons, 2000; Wilson & Schooler, 1991). Moreover, recent work shows that choice confidence can diminish when others make the same choice but provide reasons differing from one's own (Lamberton, Naylor, & Haws, 2013). In contrast, ratings are concise and easily understood, representing the evaluations of diverse peers on a scale that is commonly understood. These properties allow ratings to be surprisingly useful in real-world decision settings.

In a prominent illustration of the predictive value of ratings, Gilbert et al. (2009) asked participants to forecast their enjoyment of an experience, based on either descriptive information about the experience or the rating of another participant who had undergone the same experience. For example, in a 'speed-dating' exercise, participants were asked to forecast their enjoyment of each 'date'; as a basis for the forecast, some participants received a photograph and a descriptive profile of their partner, while others received the rating of another (unknown) participant who had already dated that partner. Results showed forecasts to be considerably more accurate when based on a simple rating than when based on descriptive information. The authors attributed these results to two phenomena: 1) systematic errors in mental simulation disrupt the use of descriptive information in making forecasts, and 2) affective reactions of different people are often surprisingly similar, especially when they belong to the same group (a consequence of homophily).

The Gilbert et al. (2009) results represent compelling evidence that the prior reactions of another individual can provide valuable information for affective forecasting. Our work builds upon this notion in the context of consumer WOM by considering the form in which such information is conveyed (ratings, commentary, or their combination.) In particular, a commentary provides not only descriptive product information, but also the reviewer's subjective opinions about the product. While this information may in fact induce errors in mental simulation, it also allows readers to make inferences regarding both the reviewer's evaluation and underlying reasons for that evaluation. To assess the relative value of rating and commentary, therefore, it is important to understand the processes by which each form of WOM is utilized by consumers to generate forecasts, and to identify factors that inhibit or facilitate each process. In the following sections, we develop a model of WOM-based forecasting in which a crucial factor is the extent to which the reviewer and prospective consumer share similar product-level preferences.

Source–receiver preference similarity

Substantial evidence indicates that consumers look for—and are persuaded by—information provided by similar peers (Forman, Ghose, & Wiesenfeld, 2008; Gershoff, Mukherjee, & Mukhopadhyay, 2007; Price, Feick, & Higie, 1989). However, in most prior research, similarity is defined in terms of group-level characteristics (gender, expertise, etc.) rather than individual-level preferences. In order to predict the relative usefulness of different WOM, we propose a conceptually distinct moderator, *source–receiver* (*S–R*) *preference similarity*, defined as the overlap in product-specific preferences of the source and receiver of WOM (e.g., a reviewer and prospective consumer); Berlo (1960) and Rothwell (2010) provide relevant communication frameworks. In principle, S–R preference similarity captures the difference in the two individuals' utility functions for a product (i.e., weighting and valuation of product attributes).

The most direct approach to measuring S–R preference similarity is to compare the actual product evaluations of source and receiver, and we utilize this approach in two studies. In the marketplace, however, actual product evaluations of prospective consumers cannot be known in advance. On the other hand, consumers (and marketers) often do know whether liking of a product varies at the aggregate level. Such knowledge is captured by the notion of *preference heterogeneity*, i.e., the extent to which preferences for a specific product or service vary within a population (Fieck & Higie, 1992; Gershoff & West, 1998; Price et al., 1989). In terms of a preference map, products with highly heterogeneous preferences (e.g., restaurants, nightclubs, paintings) are represented by a diffuse set of ideal points, while products with more homogenous preferences (e.g., mechanics, desk lamps, dry cleaners) are represented by a tightly clustered set of ideal points. Within our context, preference heterogeneity is a fundamental driver of S–R preference similarity. For products characterized by heterogeneous preferences, evaluations will differ substantially across consumers, so a prospective consumer is unlikely to encounter a reviewer with similar preferences (i.e., average levels of S–R preference similarity will be low). For products characterized by homogeneous preferences (e.g., mechanics, dry cleaners), evaluations differ little across consumers; so a prospective consumer is very likely to encounter a reviewer with similar preferences (i.e., average levels of S–R preference similarity will be high).

A model of WOM-based forecasting

We conceptualize the use of WOM in forecasting by adopting an anchoring-and-adjustment framework (Lichtenstein, Slovic, Fischhoff, Layman, & Combs, 1978; Tversky & Kahneman, 1974), as illustrated in Fig. 1. In our framework, receivers

[1] In a pretest, 189 undergraduate students were asked how helpful both a rating and a commentary would be for predicting one's enjoyment of a movie (1 = "not at all helpful," 7 = "very helpful"). Results indicated that commentary was considered more helpful (M = 3.68 vs. 5.63, $F(1, 188) = 203.03$, $p < .001$).

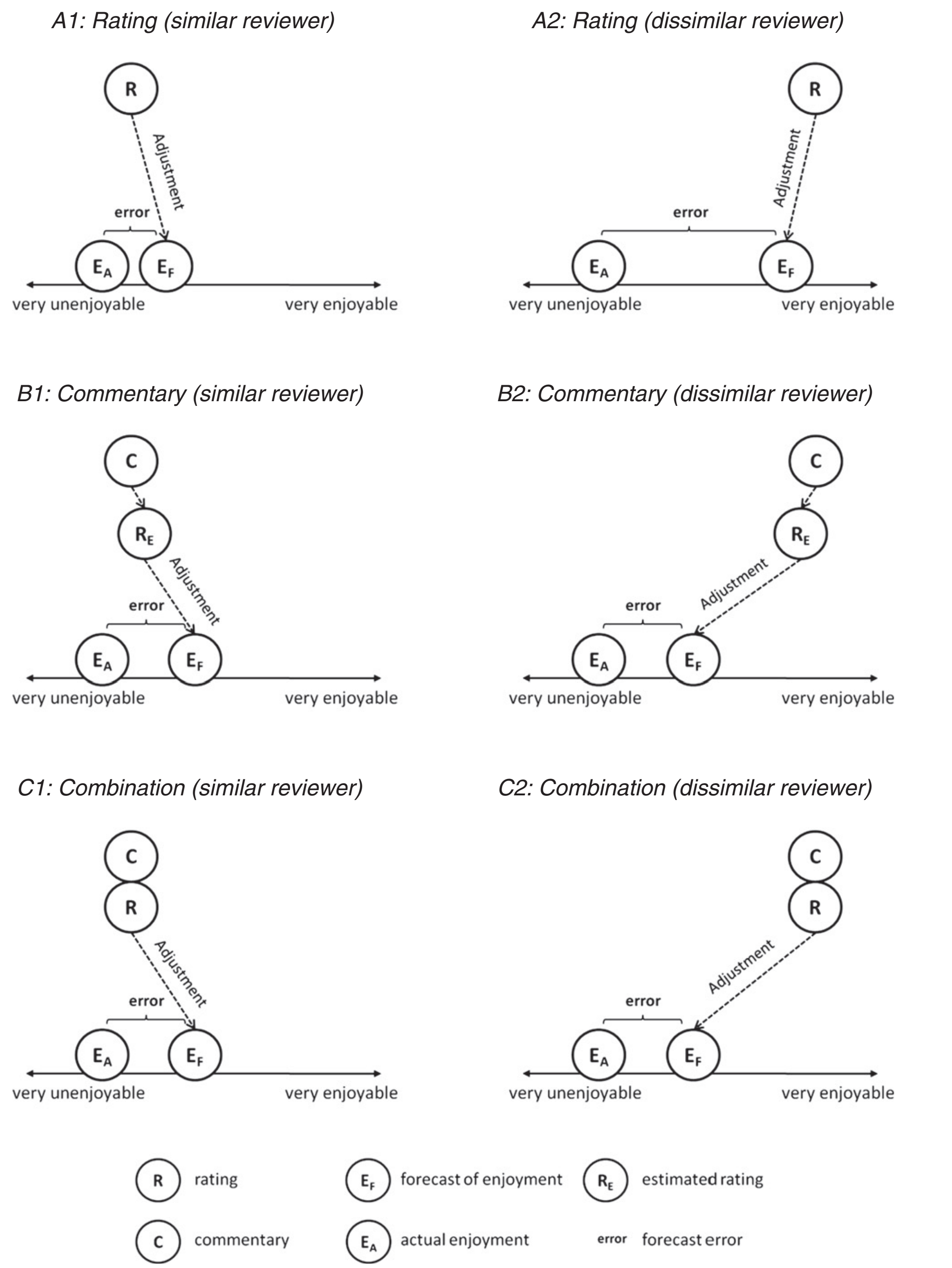

Fig. 1. Graphic interpretation of WOM-aided forecasting. A1: Rating (similar reviewer). A2: Rating (dissimilar reviewer). B1: Commentary (similar reviewer). B2: Commentary (dissimilar reviewer). C1: Combination (similar reviewer). C2: Combination (dissimilar reviewer).

estimate the evaluation of the source (reviewer), then adjust that evaluation based on the extent to which they believe their own preferences align with those of the reviewer (cf. egocentric models for predicting others' preferences — Davis et al., 1986). If WOM consists merely of an overall rating (panels A1–A2 in the figure), then the rating serves as a natural and readily available forecasting anchor, and existing research confirms that consumers often rely on others' ratings to estimate their own (Irmak, Vallen, & Sen, 2010). Even if preference similarity is unknown, consumers may adjust their predictions from this anchor: e.g., extremity aversion may provoke an adjustment towards neutrality, optimism or pessimism may provoke adjustment upward or downward, and prior experience in the product category may provoke adjustment consistent with that experience. Nonetheless, our model assumes that the extent of any adjustment is typically small. This assumption is consistent with the "assumed similarity" principle of social cognition (Cronbach, 1955), as well as recent evidence in consumer research. In particular, Naylor et al. (2011) have shown that consumers tend to perceive themselves as highly similar to an ambiguous information source, whether or not such perceptions are warranted.

As a result, we expect that when forecasts are based on rating alone: a) minimal adjustment will occur, and b) any adjustment that does occur will be of limited value. Therefore, error will be minimized when S–R preference similarity is high, and error will be maximized when S–R preference similarity is low. For example, assume that a reviewer assigns a high rating to an apartment complex, based in large part on the quality of its pool facilities. Prospective renters who encounter this WOM will tend to adjust their forecasts minimally from the high rating provided by the reviewer. As a result, forecast error should be greater for someone who does not swim, due to the dissimilarity in source and receiver preferences for this attribute.

In contrast, we suggest that forecasts based on commentary are less dependent on preference similarity. When WOM consists only of a commentary (panels B1–B2 of the figure), consumers make an estimate of the reviewer's rating as their anchor, and then use similarity cues in the commentary to adjust that anchor. Although a commentary lacks a direct indication of the reviewer's evaluation, it provides descriptive semantic content conducive to visualization and mental simulation (Gilbert & Wilson, 2007; Kahneman & Tversky, 1982), which are not dependent on similarity. By use of this content, readers are able not only to form an estimate of the reviewer's evaluation, but also to infer the reasons for that evaluation, and thereby contrast the reviewer's preferences with their own. Following the example above, assume that a review commentary speaks favorably of an apartment complex, highlighting the quality of its pool facilities. Upon encountering that commentary, a prospective renter who does not swim may be expected to: 1) perceive the reviewer's positive overall evaluation, 2) recognize the impact of the pool on this evaluation, and 3) adjust her own forecast downward. Because S–R preference similarity is identified and adjusted for, its impact is reduced.

Note that our interaction model does not predict a general superiority of commentary over ratings. Accounting for interpersonal differences is inherently difficult (Davis, Hoch, & Ragsdale, 1986; Hoch, 1988), and as discussed above, estimates based on mental simulation are subject to flaws of misinterpretation, egocentric bias, focalism, etc. (Wilson et al., 2000). Therefore, if a commentary induces consumers to make greater adjustment from the perceived reviewer anchor, such adjustment may or may not reduce forecast error, depending on S–R preference similarity. On the one hand, consumers might adjust their forecast in the wrong direction from the anchor; on the other hand, they might over-adjust, by updating their forecast too far in the proper direction. When the preferences of reviewer and reader are highly dissimilar, these concerns should be negligible compared to the benefits of commentary for adjustment, and forecasts based on commentary will outperform those based on ratings. However, when reviewer and reader have similar preferences, the 'natural anchor' of a rating is useful by itself for prediction, and the inherent flaws of commentary will often outweigh its benefits for adjustment.

Review platforms often provide rating and commentary information together. In this case, consumers receive not only an "error-free" anchor of the reviewer's evaluation, but also an underlying commentary that can be used to make similarity-based adjustment (panels C1–C2 in the figure). Although intuition suggests a synergy by which the combination is more useful than either rating or commentary alone, we argue that this synergy need not be obtained. Based on abundant evidence that individuals tend to overweight vivid or case-based information compared to statistical or numeric information (Borgida & Nisbett, 1977; Dickson, 1982; Schlosser, 2011), we expect that consumers given combined WOM will rely heavily on the commentary in making their forecast. Therefore, the provision of a commentary without a rating should invoke similar processing patterns and similar levels of forecast accuracy. In particular, forecasts based on combined information will still be subject to the errors of interpretation and simulation described above. When consumer and reviewer have dissimilar preferences, these errors are trivial compared to the benefits of commentary for adjustment, but as preferences become more similar, the value of adjustment diminishes.

Our discussion thus far has been restricted to WOM from a single reviewer. However, review platforms often provide the average rating of all reviewers, and both consumers and marketers may assume these average ratings to be especially valuable for forecasting. This intuition is consistent with the notion of the "wisdom of crowds," by which averaged group judgments are more accurate than judgments of individuals within the group (Gigone & Hastie, 1997; Larrick & Soll, 2006). However, unlike the objective judgments shown to benefit from aggregation, product preferences are inherently idiosyncratic. Therefore, the usefulness of an average product rating for forecasting should depend on the dispersion of those preferences. When preferences are highly disperse, S–R preference similarity between the 'average' reviewer and a prospective consumer will tend to be low, so that the average rating provides a poor anchor for predicting one's own evaluation. However, when preferences are more homogeneous, S–R preference similarity between the 'average' reviewer and a prospective consumer will tend to be high, so that the average rating provides a more useful anchor for prediction.

The major predictions of our framework are summarized by the following hypotheses:

H1a. The effect of WOM format on forecast error depends on S–R preference similarity. When S–R preference similarity is low, forecast error is greater for ratings (individual or averaged) than for commentary (alone or with a rating). As preference similarity increases, the difference in forecasting error between ratings and commentary diminishes.

H1b. The interaction of S–R preference similarity and WOM format is mediated by adjustment, such that 1) consumers make more adjustment when commentary is available, and 2) S–R preference similarity moderates the influence of adjustment on forecast error.

Overview of studies

All four of our studies utilized a matched-pair paradigm (Gilbert et al., 2009), in which participant 'receivers' (readers) were assigned randomly to 'sources' (reviewers) from a preliminary session. Each of the studies had three components: 1) collection of WOM from preliminary reviewers who underwent the consumption experience, 2) construction of forecasts by readers who received that WOM, and 3) actual evaluations of the consumption experience by the same readers. All four studies utilized hedonic stimuli, based on evidence that compared to utilitarian products, hedonic products are harder to quantify, more difficult to describe, and associated with lower forecasting accuracy (Huang, Lurie, & Mitra, 2009; Patrick et al., 2007; Wang et al., 2009). To ensure that participants relied solely on the WOM itself, only sparse descriptive information was presented (Gershoff, Broniarczyk, & West, 2001). Key independent variables included WOM format (ratings, commentary, or their combination) and S–R preference similarity (measured or manipulated). The primary dependent variable was forecast error, defined as the absolute difference between forecasts and evaluations.

Researchers have long known that elicitation of forecasts can impact actual experience (Olshavsky & Miller, 1972; Shiv & Huber, 2000), and that expectations may influence evaluations through elation or disappointment effects (Mellers, Schwartz, Ho, & Ritov, 1997). Therefore, it is critical to meaningfully separate forecast and evaluation, despite the challenges involved (Loewenstein & Schkade, 1999). As described below, our designs utilized multiple strategies to establish the independence of forecasting from evaluation.

Study 1

Study 1 examined the influence of different types of WOM on forecast error at different levels of S–R preference similarity. Participants in the study were asked to predict their enjoyment of different jellybeans based on the WOM of reviewers. Preference similarity was manipulated by including flavors pretested to elicit homogeneous or heterogeneous preferences. Three weeks later, participants consumed the jellybeans, and their forecasts were compared to actual enjoyment.

Method

Prior to the main study, eight different flavors of jellybean were pretested by 23 students at a large university. Participants sampled each jellybean, rated it on a 100-pt. scale (*very unenjoyable* to *very enjoyable*), and wrote a short review commentary (roughly 3–4 sentences long). These pairs of ratings and commentaries formed the collection of WOM used in the main study (Table 1 provides a sample). Based on data from the preliminary session, two flavors—cinnamon and vanilla—were chosen to manipulate preference similarity; these flavors evoked similar mean preferences but distinct variances (cinnamon vs. vanilla: M = 55.35 vs. 55.48, $F(1, 44) < 1$, NS; SD = 28.92 vs. 20.84, $F(22, 22) = 1.93$, $p = .07$). Of the other six flavors, two were chosen for use as fillers (root beer: M = 55.83, SD = 24.80; pear: M = 48.96, SD = 29.45), to reduce the likelihood that participants would associate the forecast and evaluation tasks.

Table 1
Sample ratings and commentaries.

Flavor	Rating	Commentary
Root beer	23	It's approaching (or might even be) the taste of licorice, which is a flavor I'm not a fan of. The darkness of the flavor seems to linger on my tongue, long after I'm done with it. Not a fan.
Cinnamon	86	This jellybean had a slightly hot quality to it, but in my opinion it could be hotter. It had a nice burst of flavor.
Pear	35	The appearance of the jellybean made me skeptical about its flavor. It wasn't quite as bad as I was expecting, but I would not recommend this one to my friends.
Vanilla	64	This jellybean is enjoyable. I would say that it most resembles a marshmallow sort of flavor. This makes it very enjoyable because marshmellows have a great taste. Most people who enjoy marshmellows would enjoy this flavor a lot.

One-hundred and eighteen students from the same university participated in the main study in exchange for course credit. For each of the four jellybeans (one at a time), participants were asked to read the WOM collected during the pretest, then forecast how much they would enjoy the jellybean on the same 100-pt. enjoyment scale used in the pretest. The study constituted a 2 (preference similarity: *low* vs. *high*) × 4 (WOM type: *rating* vs. *commentary* vs. *combination* vs. *avg. rating*) mixed design. As described above, preference similarity was manipulated within-subjects by use of two flavors (cinnamon and vanilla). WOM type was manipulated between-subjects following a randomized-pair approach common in social prediction research (Dunning, Griffin, Milojkovic, & Ross, 1990; Gilbert et al., 2009): In the *rating* condition, each participant viewed one rating, randomly chosen, from those collected in the pretest; in the *commentary* condition, each participant viewed one commentary; in the *combination* condition, each participant viewed both rating and commentary (from the same reviewer); and in the *avg. rating* condition, each participant viewed the average rating of the pretest group. With the exception of the *avg. rating* condition, the WOM provided to a participant for each jellybean was generated by a different reviewer, and randomization was constrained to ensure that ratings and commentaries from each reviewer were presented at an approximately equal rate. In addition to making their forecasts, participants answered two process questions (*forecast confidence* and *perceived reviewer enjoyment*, described below).

Approximately three weeks later, participants were invited back for the evaluation stage of the study. All participants tasted the four jellybeans, in an order different from that used in the forecasting stage; study materials made no mention of the prior session. Participants reported how much they enjoyed each jellybean on a 100-pt. enjoyment scale.

Forecast error

For each jellybean, forecast error was operationalized as the absolute difference between a participant's forecast and evaluation. Therefore, participants exhibiting lower forecast error were more accurate in predicting their subsequent evaluations.

Forecast confidence

After making each forecast, participants reported their confidence in that forecast on a 7-pt. scale (1 = "not at all confident," 7 = "very confident").

Perceived reviewer enjoyment and Adjustment

For each jellybean, participants were asked to indicate the reviewer's rating on the 100-pt. scale. For the *rating*, *combination*, and *avg. rating* conditions, this perceived enjoyment measure verifies that the rating was encoded accurately; in the *commentary* condition, perceived enjoyment captures participants' estimates of the reviewer's evaluation. Adjustment was calculated as the absolute difference between perceived reviewer enjoyment and a participant's own forecast. Therefore, a large adjustment indicates that a participant consciously expected his or her own evaluation to differ from that of the reviewer.

Results and discussion

As a manipulation check, we first compared S–R preference similarity for the low-similarity stimulus (cinnamon) and high-similarity stimulus (vanilla). Preference similarity was computed by subtracting from 100 the absolute difference between the evaluation of each participant and reviewer. Confirming the success of the manipulation, average preference similarity was lower for cinnamon than for vanilla ($M = 67.16$ vs. 75.14, $F(1, 228) = 7.94$, $p < .01$).

Mean forecast errors are plotted in Fig. 2, and Table 2 summarizes forecast error for all studies. H1a was tested by using a mixed-effect model to predict forecast error as a function of WOM type, preference similarity and their interaction. Analyses revealed a main effect for WOM type ($F(3, 228) = 3.14$, $p = .03$), but not for similarity ($F(3, 228) < .01$, NS). More importantly, and consistent with predictions, analyses revealed a significant interaction between WOM type and preference similarity ($F(3, 228) = 6.77$, $p < .001$), as well as the hypothesized partial interactions (*commentary* vs. *rating* $F(1, 228) = 6.79$, $p = .01$; *combination* vs. *rating* $F(1, 228) = 7.34$, $p < .01$; *commentary* vs. *avg. rating* $F(1, 228) = 12.48$, $p < .001$; *combination* vs. *avg. rating* $F(1, 228) = 13.13$, $p < .001$). Furthermore, analysis showed no support for a *commentary* vs. *combination* partial interaction ($F(1, 228) < 1$, NS; this finding was replicated in studies 2a–2b).

Follow-up comparisons revealed a pattern consistent with H1a. Under low preference similarity (cinnamon), forecast error was larger in the *rating* condition ($M = 30.93$) than both the *commentary* condition ($M = 20.22$; $F(1, 228) = 5.12$, $p = .03$) and the *combination* condition ($M = 12.37$; $F(1, 228) = 14.94$, $p < .001$). Similarly, forecast error in the *average rating* condition ($M = 28.76$) was not significantly different from error in the *rating* condition ($F(1, 228) = .20$, NS), but was greater than error in the *commentary* and *combination* conditions ($F(1, 228) = 3.39$, $p = .07$; $F(1, 228) = 12.09$, $p = .001$). However, under high preference similarity (vanilla), the advantage of commentary was eliminated: forecast error in the *rating* condition ($M = 23.63$) was not reliably different from that in the *commentary* condition ($M = 30.00$; $F(1, 228) = 1.97$, $p = .16$) or the *combination* condition ($M = 23.10$; $F(1, 228) < 1$). Forecast error in the *average rating* condition ($M = 15.83$) was not significantly different from that in the *rating* condition ($F(1, 228) = 2.82$, $p = .10$) or the *combination* condition ($F(1, 228) = 2.59$, $p = .11$), though it was lower than that in the *commentary* condition ($F(1, 228) = 10.14$, $p = .01$). In sum, the provision of commentary led to lower forecast error only when the source and receiver had dissimilar preferences; moreover, when preferences were dissimilar, the average rating resulted in more forecast error than commentary from a single reviewer.

Table 3 summarizes mean adjustment by WOM type for all studies. A mixed-effect model was used to predict adjustment as a function of WOM type, preference similarity, and their interactions. Analysis identified main effects for WOM type ($F(3, 228) = 8.02$, $p < .001$) and preference similarity ($F(1, 228) =$

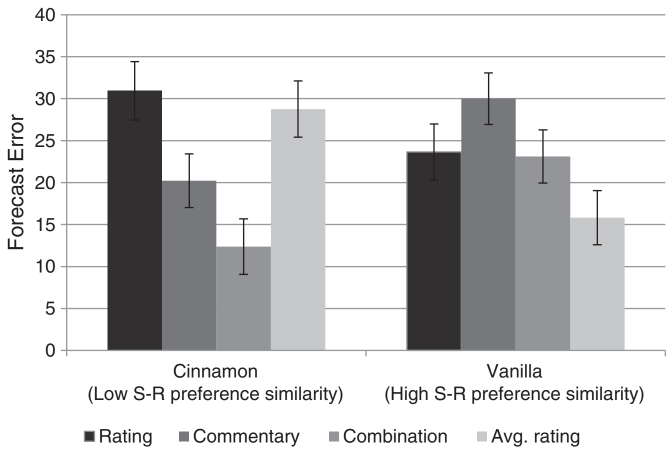

Fig. 2. Study 1: Forecast error by WOM type and S–R preference similarity. Notes: Forecast error measures the absolute value of the difference between prediction and subsequent evaluation.

Table 2
Study 1: Forecast error by WOM type and S–R preference similarity.

Study	Condition	Rating	Commentary	Combination	Avg. rating
1	Low similarity (cinnamon)	30.93	20.22	12.37	28.76
		(3.48)	(3.20)	(3.31)	(3.36)
	High similarity (vanilla)	23.63	30.00	23.10	15.83
		(3.34)	(3.07)	(3.17)	(3.22)
2a	Low similarity (−1 SD)	31.91	18.65	23.60	n/a
		(1.80)	(1.63)	(1.50)	
	Avg. similarity	20.48	16.78	18.87	n/a
		(1.15)	(1.12)	(1.14)	
	High similarity (+1 SD)	9.05	14.91	14.13	n/a
		(1.56)	(1.64)	(1.64)	
2b	Low similarity	25.63	22.31	19.95	n/a
		(2.04)	(1.98)	(1.82)	
	High similarity	11.48	17.58	16.63	n/a
		(1.67)	(1.86)	(2.09)	
3 [a]	Low similarity (−1 SD)	39.93	28.22	32.61	n/a
		(2.60)	(2.43)	(2.70)	
	Avg. similarity	26.98	23.22	23.01	n/a
		(2.00)	(2.04)	(2.04)	
	High similarity (+1 SD)	13.13	17.88	12.74	n/a
		(2.89)	(2.92)	(3.26)	

Notes: Standard errors are reported in parentheses. Lower forecast error indicates higher accuracy in predicting subsequent evaluations (and thus more useful WOM).

[a] The estimated forecast error means of study 3 were computed using *uninformed* conditions only.

8.06, $p < .01$). As expected, follow-up comparisons revealed that adjustment was significantly lower in the *rating* condition (M = 14.02) than in the *commentary* or *combination* conditions (M = 20.98, $F(1, 228) = 4.67$, $p = .03$; M = 20.35, $F(1, 228) = 3.75$, $p = .05$). Moreover, adjustment in the *avg. rating* condition was even lower than that in the *rating* condition (M = 7.31; $F(1, 228) = 4.14$, $p = .04$), suggesting that participants were especially likely to conform to aggregate opinions (Watts & Dodds, 2007).

Table 3
Studies 1–3: Adjustment by WOM type.

Study	WOM type		n	Adjustment M	Adjustment SD
1	Rating		54	14.02	2.37
	Commentary		64	20.98	2.18
	Combination		60	20.35	2.25
	Avg. rating		58	7.31	2.29
2a	Rating		158	11.87	1.29
	Commentary		162	23.98	1.25
	Combination		157	23.17	1.28
2b	Rating		132	13.96	1.29
	Commentary		120	19.01	1.33
	Combination		117	16.71	1.36
3	Rating	Uninformed	80	12.58	2.31
		Informed-similar	72	9.43	2.44
		Informed-dissimilar	72	28.66	2.45
	Commentary	Uninformed	76	26.82	2.37
		Informed-similar	74	22.25	2.41
		Informed-dissimilar	74	29.83	2.41
	Combination	Uninformed	76	19.14	2.37
		Informed-similar	70	18.59	2.47
		Informed-dissimilar	70	29.22	2.48

Notes: Adjustment was measured by comparing participants' own forecasts to their indication of the rating assigned by the reviewer. A higher score indicates greater adjustment (range: 0 to 100).

To examine the process by which S–R preference similarity moderates the influence of WOM type on forecast error, we followed the guidelines proposed by Muller, Judd, and Yzerbyt (2005) for testing mediated moderation. Our focus was the presence vs. absence of commentary from a single individual; therefore, we pooled the *commentary* and *combination* conditions and compared them to the *rating* condition (the *average rating* condition was excluded). Table 4 presents results for study 1 along with studies 2a–2b. In the first step, analysis of forecast error revealed a significant interaction effect of WOM type and preference similarity (equation 1; $\beta = 17.54$, $t(311) = 2.90$, $p < .01$). In the second step, analysis of adjustment revealed a significant effect of WOM type (equation 2; $\beta = 10.45$, $t(311) = 2.35$, $p < .05$). In the final step, both adjustment and the adjustment by preference similarity interaction were added as predictors to the first equation. The coefficient for the adjustment by preference similarity interaction was directional but nonsignificant (equation 3; $\beta = .25$, $t(309) = 1.70$, $p < .10$). Therefore, we observed only tentative support for the process suggested in H1b.

Analysis of forecast confidence revealed a main effect of WOM type ($F(3, 228) = 15.65$, $p < .001$). Consistent with the argument that consumers believe commentary to be useful, follow-up comparisons indicated that participants in the *commentary* and *combination* conditions had similar confidence in their forecasts (M = 5.52 vs. 5.18, $F(1, 228) = 1.78$, $p = .18$), and that both groups were more confident than participants in the *rating* condition (M = 3.93, $F(1, 228) = 38.62$, $p < .001$; $F(1, 228) = 23.42$, $p < .001$) or the *avg. rating* condition (M = 4.45, $F(1, 228) = 18.07$, $p < .001$; $F(1, 228) = 8.31$, $p < .001$). However, to the extent that participants were able to accurately gauge the usefulness of the provided WOM, their forecast confidence should show a strong negative correlation with actual forecast error. Contrary to this prediction, the correlation between confidence and forecast error was both small in magnitude and not significant ($r = -.16$, NS), and the correlation did not differ significantly across conditions ($\chi^2 = 4.46$, NS). Subsequent analyses for studies 2–3 (not reported here) revealed a similar lack of correlation between confidence and forecast error. In sum, participants showed limited ability to recognize the usefulness of the WOM provided.

Consistent with our conceptual framework, study 1 demonstrated that the impact of different WOM on forecast error depends on the extent to which source and receiver have similar preferences. Follow-up analyses supported our argument that different forecasting strategies were adopted depending on the WOM available, such that participants given commentary tended to adjust their prediction further from the reviewer's own evaluation. Moreover, participants had little insight into the value of WOM for their predictions. However, these findings are qualified by a limitation in the design of the study: because S–R preference similarity was manipulated by use of different products, it is conceivable that differences in the products themselves may have been responsible for our results. Furthermore, the study obtained only marginal support for the

Table 4
Mediated moderation analyses.

Study 1 predictors	Equation 1 (forecast error)		Equation 2 (adjustment)		Equation 3 (forecast error)	
	Beta	t(311)	Beta	t(311)	Beta	t(309)
WOM type	−14.51	−3.40 **	10.45	2.35 **	−13.31	−3.04 **
Preference similarity	−7.30	−1.45	−3.00	−0.57	−10.82	−1.99 **
WOM type × similarity	17.54	2.90 **	−7.58	−1.21	15.94	2.61 **
Adjustment					−0.11	−1.23
Adjustment × similarity					0.25	1.70 *
Study 2a predictors	**Equation 1 (forecast error)**		**Equation 2 (adjustment)**		**Equation 3 (forecast error)**	
	Beta	t(473)	Beta	t(473)	Beta	t(471)
WOM type	−32.18	−5.89 **	24.54	4.03 **	−13.87	−2.45 **
Preference similarity	−0.57	−9.61 **	−0.10	−1.53	−0.72	−12.24 **
WOM type × similarity	0.40	5.67 **	−0.17	−2.19 **	0.16	2.25 **
Adjustment					−0.94	−7.88 **
Adjustment × similarity					0.01	7.74 **
Study 2b predictors	**Equation 1 (forecast error)**		**Equation 2 (adjustment)**		**Equation 3 (forecast error)**	
	Beta	t(365)	Beta	t(365)	Beta	t(363)
WOM type	−4.78	−1.95 **	4.22	1.77 *	−3.46	−1.58
Preference similarity	−14.03	−5.30 **	−5.61	−2.18 **	−26.61	−9.85 **
WOM type × similarity	10.38	3.19 **	−0.51	−0.16	6.75	2.30 **
Adjustment					−0.31	−5.03 **
Adjustment × similarity					0.96	9.95 **

Notes: In these analyses, WOM type was recoded as a dichotomous variable reflecting the presence or absence of commentary (0 = *rating* condition, 1 = *commentary* and *combination* conditions).
* $p \leq .10$.
** $p \leq .05$.

processing model proposed in H1b. Studies 2a and 2b address these issues directly.

Study 2a

Study 1 relied on product-level heterogeneity as a proxy for S–R preference similarity, under the assumption that (on average) similarity between randomly paired reviewers and readers will be higher for products with homogenous preferences. In study 2a, we employed a design that allows preference similarity to vary within the same product. Participants in the studies were asked to predict their enjoyment of different music clips, based on the WOM of a reviewer. In an ostensibly unrelated task, they also evaluated a series of music clips that included the target stimuli. As before, forecasts were measured by comparing actual to predicted enjoyment.

For each clip, preference similarity was measured directly, by comparing the rating assigned by the participant to that assigned by the reviewer. According to our model, being 'matched' with a reviewer whose preferences are similar should improve the value of ratings substantially, but have little impact on the value of commentary. Therefore, the forecasting advantage of commentary over ratings should diminish as the ratings of participant and reviewer become more similar. In a supplementary analysis, we employed textual analysis to identify specific aspects of commentary content that facilitate or inhibit forecasting.

Method

By use of an initial pretest similar to that of study 1, three target music clips were selected as the focal consumption experience. The clips, which were shortened from their original length to 60 s, represented three distinct genres (country, jazz, rock) and were unfamiliar to pretest participants. Twenty undergraduate students listened to each clip, rated their enjoyment on a 100-pt. scale, and wrote a brief commentary. Average enjoyment for each clip was as follows: $M_{country}$ = 68.39 (SD = 19.13), M_{jazz} = 51.32 (SD = 25.32), M_{rock} = 53.33 (SD = 25.55). These ratings and commentaries formed the collection of WOM used in the main study.

The main study incorporated a 3 (WOM type: *rating* vs. *commentary* vs. *combination*) × 1 (S–R preference similarity) × 3 (music genre: *country* vs. *jazz* vs. *rock*) mixed design. S–R preference similarity was a continuous measured variable, described below. The WOM type manipulation was identical to that of study 1, and music genre was treated as a control variable in the analysis.

One-hundred sixty-five students from the same university participated in the study for course credit. The study consisted of two consecutive phases. In the first phase, participants were presented with WOM (rating, commentary, or combination, depending on condition) for each of four different music clips. The first of these was an unrelated filler, followed by WOM for the three target clips; presentation of WOM followed the same constrained randomization as study 1. After reviewing each WOM, participants forecasted how much they would enjoy that music on a 100-pt. scale. They also provided measures of forecast confidence and perceived reviewer enjoyment (identical to those in study 1).

After making their forecasts, participants were told that their next task was an unrelated pretest of various pieces of music. All participants then listened to four clips: the first was a decoy clip not relevant to the study, and the following three were the target clips, presented in an order different from the forecast stage. Participants reported how much they enjoyed each music clip on the same 100-pt. enjoyment scale. Finally, they rated their liking of various musical genres, including the three target genres, on a 7-pt. scale (−3 = "dislike very much," 3 = "like very much").

Results and discussion

The S–R preference similarity variable was constructed as before, by subtracting from 100 the absolute difference between the evaluation of each participant and reviewer. H1a was tested by using a mixed-effect model to predict forecast error as a function of WOM type, S–R preference similarity, music genre, and their interactions. Analyses revealed main effects for WOM type ($F(2, 459) = 17.89$, $p < .001$) and preference similarity ($F(1, 459) = 79.54$, $p < .001$). More important, and consistent with our hypothesis, analyses revealed a significant overall interaction between WOM type and preference similarity ($F(2, 459) = 16.45$, $p < .001$), as well as the hypothesized partial interactions (*commentary* vs. *rating* $F(1, 459) = 31.26$, $p < .001$; *combination* vs. *rating* $F(1, 459) = 16.78$, $p < .001$; see Table 2 and Fig. 3).

To explore the nature of the overall interaction, we examined the effects of WOM format on forecast error at low and high levels of preference similarity by use of spotlight analysis (Irwin & McClelland, 2001). Consistent with predictions, planned contrasts at low levels of preference similarity (one SD below the mean) revealed that forecast error was significantly smaller for both the *commentary* condition ($M = 18.65$) and the *combined* condition ($M = 23.60$) than for the *rating* condition ($M = 31.91$, $t = 5.46$, $p < .001$; $t = 3.55$, $p < .001$). As expected, however, planned contrasts at high levels of preference similarity (one SD above the mean) revealed an opposite pattern: forecast error was significantly larger for both the *commentary* condition ($M = 14.91$) and *combined* condition ($M = 14.13$) than for the *rating* condition ($M = 9.05$, $t = 2.59$, $p = .01$; $t = 2.25$, $p = .03$).[2]

Next, the adjustment measure was examined with a mixed-effect model including WOM type, preference similarity, and their interactions (see Table 3 for means). Analysis identified main effects for WOM type ($F(2, 459) = 9.13$, $p < .001$), preference similarity ($F(1, 459) = 31.38$, $p < .001$), and their interaction ($F(2, 459) = 4.39$, $p = .01$). As before, adjustment in the *rating* condition was modest overall ($M = 11.87$), supporting our argument that rating-based forecasts tend to invoke only limited adjustment. Furthermore, adjustment in the *rating* condition was significantly lower than that in both the *commentary* and *combination* conditions ($M = 23.98$, $F(1, 459) = 45.27$, $p < .001$; $M = 23.17$, $F(1, 459) = 38.65$, $p < .001$).

As in study 1, a test of mediated moderation was conducted to examine the underlying process proposed in H1b. Results are depicted in Table 4. The first step revealed a significant interaction effect of WOM type and preference similarity on forecast error (equation 1; $\beta = .40$, $t(473) = 5.67$, $p < .001$). The second step revealed a significant effect of WOM type on adjustment (equation 2; $\beta = 24.54$, $t(473) = 4.03$, $p < .001$). Finally, the third step revealed a significant interaction effect of adjustment and preference similarity on forecast error, after controlling for the predictors in step 1 (equation 3; $\beta = .01$, $t(471) = 7.74$, $p < .001$). Next, the bootstrapping procedure recommended by Hayes (2012) was performed at both low (−1 SD) and high (+1 SD) levels of preference similarity. In both cases, the 95% confidence interval for the indirect effect did not contain zero (95% CI at low similarity = [−4.90, −1.64]; 95% CI at high similarity = [1.47, 5.27]). Together, these findings support H1b and our argument that: 1) the presence of commentary in WOM increases adjustment, but 2) this adjustment reduces forecast error only when preference similarity is low.

Findings of study 2a replicated those of study 1 by showing that S–R preference similarity (here measured directly) influences the relative value of different WOM types for consumer forecasting. When participants were matched with reviewers having very different preferences, commentary produced the lowest forecast error, but when reviewers had very similar preferences, rating produced the lowest forecast error. Furthermore, analysis of the adjustment mediator supported our argument that consumers given a rating alone adjust their forecast only minimally from that rating, while consumers given commentary utilize its content to infer preference similarity and adjust their forecast accordingly. To provide deeper insight into the means by which commentary enables inferences of preference similarity, we next conducted a follow-up analysis of the commentaries themselves.

Commentary analysis

In an exploratory investigation, we examined the textual content of commentaries utilized in the first two studies. Our goal was to identify characteristics of the text that relate to: 1) estimation of the reviewer's evaluation (anchoring), and 2) inferences of similarity with the reviewer (adjustment). In the analysis, the 46 commentaries from study 1 and 60 commentaries from study 2a were assessed individually, using the Linguistic Inquiry and Word Count program (LIWC — Pennebaker, Booth, & Francis, 2007). LIWC is based on a matching algorithm: after receiving a target script, the program searches that script for

[2] A valid concern raised by the use of actual S–R preference similarity is the fact that the receiver's evaluation must be known a priori. Subsequent analysis showed that the same pattern of results is obtained when music genre preferences of participants were used in place of their actual evaluations. For each of the three clips, we compared the reviewer's evaluation of the clip (transformed to a 7-pt. scale) to the participant's evaluation of the genre as a whole, then rescaled the difference so that a higher number indicated greater similarity. Consistent with the findings above, results of a mixed-effect analysis revealed a significant overall interaction of WOM type and genre-based preference similarity ($F(2, 459) = 5.47$, $p < .01$), with means that followed the pattern described above. In addition, the expected partial interactions remained significant: *commentary* vs. *rating* $F(1, 459) = 8.25$, $p < .01$; *combination* vs. *rating* ($F(1, 459) = 8.39$, $p < .01$).

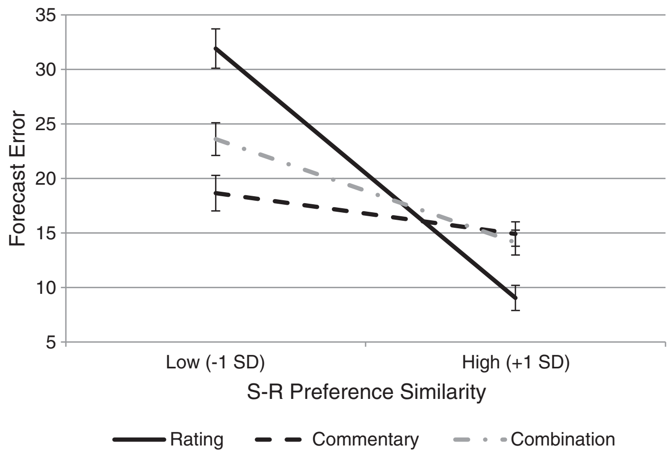

Fig. 3. Study 2a: Forecast error by WOM type and S–R preference similarity. Notes: Forecast error measures the absolute value of the difference between prediction and subsequent evaluation.

words representing 70 linguistic or psychological dimensions, then calculates the percentage of total words in the script that fall into these dimensions. LIWC has been extensively validated and applied to a wide variety of substantive domains, including physical and psychological health, interpersonal relationships, and deception (Campbell & Pennebaker, 2003; Newman et al., 2003).

In order to investigate the influence of linguistic components on estimation, we restricted our examination to the *commentary* condition (which did not receive the reviewer's rating directly). Our inquiry took place in three steps. First, *estimation error* was defined as the absolute difference between perceived reviewer enjoyment and actual reviewer enjoyment, and an average estimation error was calculated for each of the 106 commentaries. Next, each of the commentaries was submitted to LIWC and assigned a score on each underlying dimension. Finally, correlation analyses were conducted to identify linguistic or psychological dimensions of the commentaries that predicted their average estimation error.

Analyses revealed that, on average, longer reviews did not enhance estimation ($r = .01$, $p = .91$). However, estimation error was associated with two theoretically relevant LIWC categories. First, estimation error was smaller for commentaries that made greater use of *affect* words (e.g., 'enjoy,' 'great,' 'awful'; $r = -.23$, $p = .02$). Given that such terms involve the direct expression of feelings, readers may logically use their valence and frequency as indicators of the reviewer's overall evaluation. Second, estimation error was smaller for commentaries that made greater use of *exclusive* words (e.g., 'lack,' 'really,' 'just'; $r = -.17$, $p = .07$). Consistent with previous arguments by Pennebaker and King (1999) that exclusive words help readers to distinguish between possible interpretations of an author's intended meaning, the presence of exclusive words may reduce ambiguity when inferring reviewers' opinions.

In order to investigate the adjustment process, we restricted our examination to participants in the *combination* condition; because these participants received the reviewer's rating, their forecast error provides a direct reflection of inaccurate adjustment. In a manner similar to that above, we first calculated the average forecast error for each of the 106 commentaries, then conducted correlation analyses to identify dimensions of the commentaries that predicted their average forecast error. Analyses revealed that, on average, adjustment error was not associated with the overall length of a commentary ($r = .06$, $p = .54$). However, adjustment error was associated with two LIWC dimensions that are both theoretically relevant and distinct from those identified above. First, adjustment error was larger for commentaries making greater use of function words (adverbs, pronouns, articles, prepositions, etc.; $r = .19$, $p = .05$). Function words have been described as 'glue' that holds more substantive content together and helps writers to clarify their opinions (Pennebaker et al., 2003). In a product review, however, greater use of function words necessarily reduces the proportion of content devoted to product- or context-relevant information, which is more useful to readers in gauging similarity with the reviewer. Second, adjustment error was larger for commentaries making greater use of the past tense ($r = .24$, $p = .01$), but smaller for commentaries making greater use of the future tense ($r = -.21$, $p = .03$). Closer inspection revealed that past tense was often used by reviewers to describe experience with the product objectively, which provides little guidance regarding similarity. In contrast, future tense was often used by reviewers to convey intentions or provide contexts in which they might consume the product. To the extent that readers can or cannot identify with these usage contexts, they may infer more or less similarity with the reviewer.

Study 2b

Although study 2a provided a direct test of our framework, its design was constrained by the use of a post hoc measure for S–R preference similarity. In addition, despite precautions, it is possible that some participants linked the WOM they had received to the clips at the evaluation stage. Study 2b addresses these concerns with a design that both: 1) manipulates S–R preference similarity directly, and 2) clearly separates the forecast and evaluation stages.

The procedure of study 2a was modified by reversing the order of prediction and evaluation. With this change, S–R preference similarity could be manipulated a priori, and potential dependencies between prediction and evaluation were minimized. Because prediction took place subsequent to evaluation, it was not a forecast in the traditional sense; however, in this study (and all others), the prediction question did not specify when consumption would occur. The order of prediction and evaluation is irrelevant if one assumes that underlying preferences do not change systematically over the interim. We believe this assumption to be reasonable for music clips and use the term *forecast* to maintain consistency.

Method

One-hundred twenty-three students from a large university were recruited to participate in a two-session, computer-based study for course credit. Stimuli (music clips) were identical to those of study 2a, and the same set of WOM information was utilized. However, the order of forecast and evaluation tasks was reversed, so that evaluation preceded forecasting. Therefore, any expectation or demand effects generated by the act of forecasting could not have influenced evaluations. In addition, a time interval of approximately three weeks was introduced between the two stages.

Because evaluation measures were collected during the first session, S–R preference similarity could be manipulated directly. Hence, the study incorporated a 3 (WOM type: *rating* vs. *commentary* vs. *combination*) × (preference similarity: *high* vs. *low*) × 3 (music genre: *country* vs. *jazz* vs. *rock*) mixed design. Participants were randomly assigned to one of the six WOM type × similarity conditions, and music genre was a within-subjects factor. The WOM type manipulation was identical to that of study 2a. Preference similarity was manipulated as follows: for each participant and music clip, actual similarity with each potential reviewer was calculated using the same method as study 2a. Next, participants in the *high-similarity* (*low-similarity*) condition were randomly paired with reviewers who had provided similar (dissimilar) ratings of the clip; the process was constrained so that WOM from each reviewer was presented an equal number of times. As intended, this procedure resulted in substantial differences in preference similarity across conditions: *high-similarity* $M = 96.54$, *low-similarity* $M = 63.24$ ($F(1, 351) = 1623.30$, $p < .001$). Finally, forecast confidence and perceived reviewer enjoyment were measured as before.

Results and discussion

Forecast errors are shown in Table 2. A mixed-effect model was used to predict forecast error as a function of WOM type, S–R preference similarity, music genre, and their interactions. Analyses revealed a main effect for similarity ($F(1, 351) = 22.40$, $p < .001$) but no main effect for WOM type ($F(2, 351) < 1$). More importantly, and consistent with hypotheses, a significant interaction indicated that the impact of WOM type on forecast error was moderated by similarity ($F(2, 351) = 4.85$, $p < .01$), and both planned partial interaction contrasts were also significant (*commentary* vs. *rating*: $F(1, 351) = 6.20$, $p = .01$; *combination* vs. *rating*: $F(1, 351) = 8.03$, $p < .01$). For participants matched with low-similarity reviewers, mean forecast error in the *commentary* condition ($M = 22.31$) and *combination* condition ($M = 19.95$) was lower than that in the *rating* condition ($M = 25.63$), but the difference was reliable only for the latter ($F(1, 351) = 1.36$, NS; $F(1, 351) = 4.31$, $p = .04$). For participants matched with high-similarity reviewers, this pattern reversed: mean forecast error in both the *commentary* condition ($M = 17.58$) and the *combination* condition ($M = 16.63$) was greater than that in the *rating* condition, though the difference was reliable only for the former ($M = 11.48$, $F(1, 351) = 5.98$, $p = .02$; $F(1, 351) = 3.72$, $p = .06$).

Examination of our adjustment measure again suggested that minimal adjustment occurred with ratings alone, but adjustment was more extensive when commentary was available. Table 3 shows the extent to which participants in each condition adjusted their own forecasts from their estimate of the reviewer's opinion. Replicating the prior studies, findings revealed that adjustment in the *rating* condition was significantly smaller than that in the *commentary* condition ($M = 13.96$ vs. 19.01, $F(1, 351) = 7.40$, $p < .01$). In contrast to the prior studies, the difference between adjustment in the *rating* condition and *combination* condition ($M = 16.71$) was only directional ($F(1, 351) = 2.15$, $p = .14$).

As before, a test of mediated moderation was used to examine our underlying processing model; results are depicted in Table 4. The first step revealed a significant interaction effect of WOM type and preference similarity on forecast error (equation 1; $\beta = 10.38$, $t(365) = 3.19$, $p < .01$), and the second step revealed a marginal effect of WOM type on adjustment (equation 2; $\beta = 4.22$, $t(365) = 1.77$, $p < .10$). The third step revealed a significant interaction effect of adjustment by preference similarity on forecast error, controlling for predictors in the first step (equation 3; $\beta = .96$, $t(363) = 9.95$, $p < .001$). Bootstrapping analyses (Hayes, 2012) were performed separately for the low-similarity condition and high-similarity condition; in both cases, the 95% confidence interval for the indirect effect did not contain zero (95% CI for low-similarity = [−2.91, −0.38]; 95% CI for high-similarity = [1.07, 5.34]).

Taken together, the first three studies provide convergent evidence for our argument that neither rating nor commentary has a consistent advantage over the other in aiding prediction. Instead, their relative value depends on whether the consumer is paired with a reviewer whose underlying preferences are similar. Results were consistent with our claim that consumers given a rating alone apply an 'assumed similarity' heuristic (Cronbach, 1955; Naylor et al., 2011), which is most effective when source and receiver indeed have similar preferences. On the other hand, consumers given commentary need not rely on such a heuristic, because similarity is inferred (albeit imperfectly) from the commentary itself.

Our final study was designed to test this logic more directly. In addition to reviewer WOM, some participants were also given explicit information regarding their preference similarity with the reviewer, and the accuracy of this information varied. For consumers receiving ratings alone, explicit similarity information provides a simple means of adjustment, so that forecast error should depend heavily on the accuracy of that information. In contrast, consumers receiving commentary have other cues for adjustment available, so that forecasts should be less affected by the accuracy of that information. Stated formally:

H2. The accuracy of explicit preference similarity information will affect forecasts based on ratings to a greater extent than forecasts based on commentary or combined information.

Study 3

The design of study 3 was similar to that of studies 1–2, with two important modifications. First, participants were provided not only different types of WOM, but also information regarding S–R preference similarity, which was sometimes accurate and sometimes inaccurate (see below). Second, we included a post-task introspection measure, in order to identify strategies used in processing different types of WOM.

Method

Target stimuli were the four flavors of jellybeans utilized in study 1, and the same collection of WOM was utilized. The study incorporated a mixed design with three factors. WOM type (*rating* vs. *commentary* vs. *combination*) was a between-subjects factor, manipulated as before. Indicated similarity (*informed-similar* vs. *informed-dissimilar* vs. *uninformed*; described below) and flavor (*root beer* vs. *cinnamon* vs. *pear* vs. *vanilla*) were both repeated factors.

One-hundred and eight university students were recruited to participate in a two-session, computer-based study for course credit. At the start of the first session, participants answered a series of survey questions about their liking for different flavors; this survey was used as a cover story for the subsequent similarity manipulation. Next, participants were exposed to WOM for each jellybean, one at a time, along with explicit information regarding their preference similarity with the reviewer. Specifically, participants in the *informed-similar* and *informed-dissimilar* conditions read that "Based on the information you shared with us earlier … this student's preferences for jellybeans are generally *very SIMILAR* (*very DISSIMILAR*) to yours." Participants were given the *SIMILAR* phrasing for two jellybeans and the *DISSIMILAR* phrasing for two jellybeans (the order was counterbalanced). In the *uninformed* conditions, participants were told nothing at all about their similarity to the reviewer.

As in the previous studies, participants then forecasted how much they would enjoy the jellybean, reported their confidence in that forecast, and provided an estimate of the reviewer's enjoyment. In an additional manipulation check, participants rated the degree to which they perceived their own preference to be similar to that of the reviewer, using a 100-pt. scale ("not at all similar," "very similar"). At the end of the session, participants completed an introspection measure in which they were asked to write "a few sentences" describing the process by which they made their forecasts. In the second session, which took place approximately three weeks later, participants tasted the jellybeans and reported their enjoyment (one at a time), using the same measures as before.

Results and discussion

Initial examination of the manipulation check confirmed that explicit similarity information influenced participants' perceptions of similarity. Compared to participants in the *uninformed* condition, participants in the *informed-similar* condition rated their own preferences as more similar to those of the reviewer ($M = 50.44$ vs 61.83; $F(1, 612) = 26.12$, $p < .001$), and participants in the *informed-dissimilar* rated their preferences as less similar to those of the reviewer ($M = 41.98$; $F(1, 612) = 14.37$, $p < .001$).

For each jellybean and participant–reviewer combination, actual S–R preference similarity was calculated in the same manner as studies 1–2. We first examined the *uninformed* conditions alone, which constitute a replication of the earlier studies. Analyses using the same mixed-effect model revealed a significant two-way interaction, by which the effect of WOM type on forecast error was moderated by actual similarity ($F(2, 598) = 5.22$, $p < .01$). This result was consistent with both hypothesis 1 and our earlier findings.

H2 argues that accurate vs. inaccurate information regarding preference similarity should affect forecast error to a greater extent when forecasts are based on ratings than when forecasts are based on commentary. The hypothesis was examined in two steps. In the main analysis, a mixed-effect model was used to estimate forecast error as a function of WOM type, indicated similarity, actual similarity, flavor, and their two-way and three-way interactions. Analysis identified main effects for WOM type ($F(2, 598) = 5.36$, $p < .01$), indicated similarity ($F(2, 598) = 9.74$, $p < .001$), and actual similarity ($F(1, 598) = 51.64$, $p < .001$). Most importantly, and consistent with H2, analyses revealed a significant three-way interaction between WOM type, indicated similarity, and actual similarity ($F(4, 598) = 3.16$, $p < .05$). Therefore, we examined the two-way interaction between indicated similarity and actual similarity at each WOM type; relevant data are depicted in the panels of Fig. 4.

For the *rating* condition (panel A), the interaction between indicated similarity and actual similarity was significant ($F(2, 598) = 13.64$, $p < .001$). A spotlight analysis was conducted to examine the effects of informed similarity at low and high levels of actual similarity. At low levels of actual similarity (one SD below the mean), forecast error was significantly larger for both the *uninformed* and *informed-similar* conditions ($M = 39.82$, 38.28) than for the *informed-dissimilar* conditions ($M = 25.27$, $t = 2.96$, $p < .01$; $t = 2.65$, $p < .01$). However, this pattern reversed under high S–R preference similarity: forecast error was significantly smaller for the *uninformed* and *informed-similar* conditions ($M = 12.91$, 9.74) than for the *informed-dissimilar* conditions ($M = 32.53$, $t = 3.74$, $p < .01$; $t = 4.06$, $p < .01$). In sum, participants given ratings alone made use of explicit similarity information in their forecasts, and they benefitted substantially when they were correctly informed that the reviewer had dissimilar preferences.

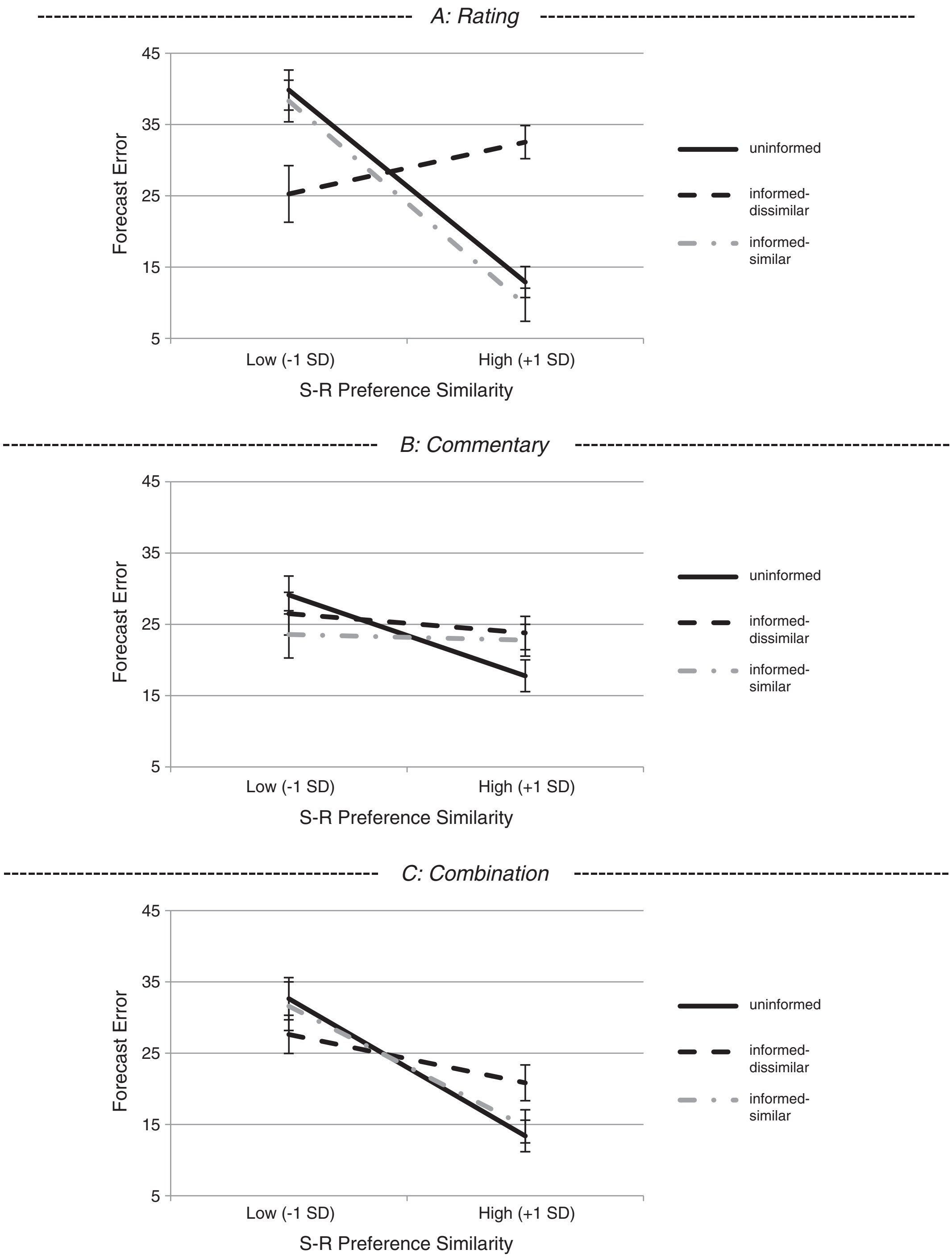

Fig. 4. Study 3: Forecast error by WOM type, indicated similarity, and S–R preference similarity.

However, the *commentary* and *combination* conditions (panels B and C of Fig. 4) showed a very different pattern. For both *commentary* and *combination*, the interaction between indicated similarity and actual similarity was not significant ($F(2, 598) = 1.09$, $p = .34$; $F(2, 598) = 1.92$, $p = .15$), nor was there a significant main effect of indicated similarity ($F(1, 598) = .90$, NS; $F(1, 598) = 1.81$, $p = .17$). Furthermore, spotlight contrasts revealed no significant affect of indicated similarity at either low or high levels of actual similarity (all Fs < 1). These findings are consistent with our claim that participants used the commentary content itself to gauge similarity with the reviewer, making explicit similarity information less useful.

In the final step, we directly compared the effect of accurate vs. inaccurate similarity information across WOM type. To do so, we calculated the difference in forecast error between *informed-similar* and *informed-dissimilar* conditions at both low levels of similarity (−1 SD) and high levels of similarity (+1 SD). At both levels of similarity, planned contrasts revealed that this difference was larger in the *rating* condition than in the *commentary* and *combination* conditions (low: $M = 12.49$, $t = 2.13$, $p < .05$; high: $M = 19.24$, $t = 2.88$, $p < .01$). Therefore, H2 was confirmed.

Means for the adjustment measure are presented in Table 3. Examination of this measure supported our argument that consumers given a rating alone will adjust little from that rating unless prompted to do so, while consumers given commentary will use its content to adjust their forecasts. First, comparison of the three *rating* groups revealed that adjustment was minimal in both the *uninformed* and *informed-similar* conditions ($M = 12.58$ and $M = 9.43$, $F(1, 628) < 1$), while adjustment in the *informed-dissimilar* condition was substantially larger ($M = 28.66$, $F(1, 628) = 22.82$, $p < .001$; $F(1, 628) = 30.95$, $p < .001$). Second, comparison of the three *uninformed* groups revealed adjustment to be significantly greater for the *commentary* and *combination* conditions ($M = 26.82$ and $M = 19.14$) than for the *rating* condition ($M = 12.58$, $F(1, 628) = 18.45$, $p < .001$; $F(1, 628) = 3.93$, $p < .05$), replicating our prior studies. Finally, explicit similarity information appeared to have little effect on adjustment when commentary was available: among the *commentary* groups, adjustment in the *uninformed* condition ($M = 26.82$) did not reliably differ from that in the *informed-similar* or *informed-dissimilar* conditions ($M = 22.25$, $M = 29.83$, NS). Among the *combination* groups, adjustment in the *uninformed* condition ($M = 19.14$) did not reliably differ from that in the *similar* condition ($M = 18.59$, NS), though it was significantly lower than that in the *informed-dissimilar* condition ($M = 29.22$, $F(1, 628) = 8.62$, $p < .01$).

Finally, participants' verbal reports provided a means of investigating the process by which forecasts were generated. The content of these reports was examined for specific words relating to the use of mental simulation (e.g., "imagine" and "taste"). Each report was coded in a binary manner for the presence or absence of such words (given that reports were typically 1–2 sentences, more complex coding schemes were not practical). A subsequent analysis of proportions revealed that reference to simulation was considerably more common in the *commentary* conditions (78%) and *combination* conditions (66%) than in the *rating* conditions (19%; $\chi^2(1) = 25.35$, $p < .001$; $\chi^2(1) = 15.57$, $p < .001$). Although preliminary, these results support our framework and identify mental simulation as a factor distinguishing the processing of commentary- and rating-based WOM.

General discussion

For the vast majority of consumer decisions, others have already experienced options under consideration and shared their own opinions. Growth in e-commerce and communications has enhanced the availability of consumer word-of-mouth, raising the question of which formats offer the greatest potential for enhancing consumer forecasts. The present research examined two common forms of WOM, numeric ratings and text commentary, and a moderating factor, S–R preference similarity. Consistent with our anchoring-and-adjustment framework, an advantage of commentary over ratings was observed for settings in which consumers encountered reviewers with dissimilar preferences. This advantage diminished when consumers encountered reviewers with similar preferences or when preference similarity information was provided directly. Furthermore, participants who received both rating and commentary together appeared to rely heavily on the commentary, resulting in similar processing patterns and similar forecast error, despite the 'added information.' Examination of underlying processing patterns revealed evidence of both mediation and moderation: i.e., the presence of commentary in WOM increases adjustment, but this adjustment reduces forecast error only when preferences are dissimilar.

Theoretical contributions

Our research contributes to a rapidly evolving literature on the multiple roles played by word-of-mouth in consumer behavior. Recent inquiries have explored factors affecting the likelihood of WOM transmission (Berger & Schwartz, 2011; Cheema & Kaikati, 2010), the type and format of WOM content (De Angelis et al., 2012; Ryu & Han, 2009; Schellekens et al., 2010), and the effects of WOM transmission on both source and receiver (Chan & Cui, 2011; Moore, 2012; Weiss et al., 2008; Zhao & Xie, 2011). However, the objective value of WOM for improving consumer decisions has received surprisingly little attention. In the area of online reviews, existing work has focused primarily on characteristics affecting persuasiveness and downstream sales (Archak et al., 2011; Chen et al., 2011; Chevalier & Mayzlin, 2006; Liu, 2006; Schlosser, 2011; Zhu & Zhang, 2010), but an emerging stream has also begun to focus on subjective helpfulness and related variables (Mudambi & Schuff, 2010; Schindler & Bickart, 2012; Sen & Lerman, 2007). We extend this discussion to focus on objective helpfulness, by examining how consumers utilize review information to generate forecasts of consumption enjoyment.

Prior research in affective forecasting has demonstrated that the rating of a single peer is often more useful for prediction than descriptive information (Gilbert et al., 2009). We supplement this idea in several ways. First, reflecting our focus on consumer WOM, we compare rating and commentary. Because both forms of WOM are filtered through the lens of the reviewer, they represent two distinct forms of 'surrogate' information whose relative value for forecasting has not been explored. Second, we propose distinct mechanisms by which each form of WOM is used for forecasting. Our model does not argue that commentary or rating is inherently superior, but rather focuses on the moderating role of S–R preference similarity. We suggest that similar to purely descriptive information, commentary trades off benefits of extra information against errors of mental simulation; however, commentary provides an additional benefit in the form of cues enabling inference of S–R preference similarity. Our studies demonstrate that the forecasting advantage of ratings over commentary is restricted to cases of high S–R preference similarity. Finally, our findings add a caveat to the notion that 'surrogation' information is useful only in homophilous environments. Although we find clear evidence for this assertion in our rating conditions, we also show that commentary—which provides both 'surrogation' information and a means of gauging its relevance—represents a valuable forecasting tool even when preferences are heterogeneous.

Substantial prior evidence indicates that consumers look for—and are more likely to be persuaded by—information from similar peers (Forman, Ghose, & Wiesenfeld, 2008; Price, Feick, & Higie, 1989). However, similarity is typically defined in terms of group-level characteristics (gender, expertise, etc.); in contrast, our key construct of S–R preference similarity captures the objective difference in preferences between source and receiver. Prior research confirms that consumers will utilize information regarding preference similarity when it is presented directly (e.g., the prior opinions of an online agent — Gershoff et al., 2003). In our studies, review commentary provided a basis for inferring similarity indirectly, and participants were at least moderately successful in doing so; our textual analysis identified features of the commentary which may have facilitated the process. In contrast, when given a rating alone, participants were remarkably willing to 'copy' that rating as their forecast, although a number of possible adjustments were feasible (moderation of extreme ratings, adjusting for category-level preferences, etc.). This tendency is consistent with the principle of "assumed similarity" (Cronbach, 1955), as well as the broader false consensus effect (Ross et al., 1977). However, our studies inverted the typical false consensus paradigm, as participants were first given another's evaluation and then asked to forecast their own. Hence, our findings contribute to recent work showing that consumers assume an ambiguous source to have similar preferences, even when this assumption is unwarranted (Naylor et al., 2011).

Practical implications

The vast majority of online retailers offer some form of review platform by which consumers may observe the feedback of their peers. Among a broad array of issues to be considered in implementing such a platform, firms must carefully consider their effects from a consumer perspective. In particular, improving the forecast accuracy of prospective consumers allows sellers to increase customer satisfaction, strengthen loyalty, and reduce return costs; therefore, it is imperative to consider the effects of WOM provision on consumer forecasting. From this view, our results challenge a number of intuitions regarding the use of ratings, reviews, and WOM more generally. Perhaps most notably, exposure to a greater quantity of WOM did not always produce more accurate forecasts. In particular, although real-world review platforms typically present rating and commentary information together, the combination conditions in our studies sometimes produced less accurate forecasts than either rating or commentary alone (studies 2–3). Moreover, across all studies, we observed low correlations between confidence and forecast error, suggesting a general lack of awareness regarding the value of WOM. Finally, our study 1 results provide tentative evidence that the provision of 'average' ratings may be of limited benefit when opinions vary greatly across consumers.

Even though S–R preference similarity cannot be known in advance at the individual level, marketers and consumers are likely to have reasonable lay theories about preference heterogeneity at the aggregate level (Gershoff & West, 1998; Price et al., 1989). In terms of product category, consumers may expect preferences to vary more widely for hedonic goods (which tend to lack agreed-upon, objective performance standards) than for functional goods, for niche products than for mass-market products, etc. In addition, consumers may also infer heterogeneity based on distributions of product ratings, prior experience with similar products, etc. Our findings suggest that the potential advantages of collecting and providing reviewer commentaries will be most pronounced when preference heterogeneity is either high or unknown. In these cases, retailers might emphasize the availability of reviewer commentary and directly encourage its use. However, the context in which reviews are encountered must also be considered: e.g., if reviews are posted on a website used by a homogeneous population, consumers may know that reviewers are likely to have similar preferences, even for products characterized by high preference heterogeneity (for an example involving music, see Naylor et al., 2011). More generally, based on customer-level data (purchase histories, customer profiles, past feedback, etc.), firms may be approximate the S-R preference similarity of a prospective customer with previous customers. Among other benefits, doing so would enable the provision of 'customized' WOM that prioritizes reviewers with similar preferences (e.g., by arranging reviews in order of 'similarity').

Although our experiments utilized a specific review context, the underlying insights apply to a variety of contemporary WOM environments. From the perspective of our anchoring-and-adjustment model, the most 'helpful' review is one that both transmits an evaluation clearly and provides sufficient cues by which readers may accurately infer similarity; more generally, consumers benefit from knowing whether their preferences are similar to those of the reviewers they encounter. Thus we offer three general principles for applying our ideas to consumer WOM more broadly. First, both the average level of S–R preference similarity and the ability of receivers to gauge that similarity will vary substantially across channels. For channels characterized by stronger ties between sender and receiver (e.g., text messaging, social networking 'circles'), receivers will usually have knowledge of their similarity with the sender, and our findings suggest that adjustment will be fairly accurate even without commentary. However, for channels characterized by relatively weaker ties between sender and receiver (e.g., blogs, discussion forums), preference similarity may be unknown, and commentary provides receivers a valuable tool for adjustment. Second, the specific form in which ratings and commentary are communicated will vary according to the channel in which they are conveyed. For example, summary evaluations may be communicated in verbal or symbolic means (e.g., a '2-star' movie evaluation may be encoded in the text message "Total waste of time! ☹"). Therefore, the evaluative 'anchor' in our model will be estimated with varying levels of precision. Finally, different channels impose different restrictions on message length or format which directly impact the value of commentary. For example, when message length is limited (e.g., 240 characters on Twitter), characteristics associated with better estimation and adjustment become especially important; our lexical analysis provides initial guidance in this regard.

Limitations and future research

Our set of studies focused on the transmission of WOM from a single source. The acquisition and aggregation of multi-sourced WOM are important topics unto themselves, and although our aggregate, 'average rating' conditions shed some initial light on this topic, further investigation is warranted. More generally, a clear need exists for the establishment of a broader model to capture exposure, attention, and integration of multiple types of WOM from multiple providers. Such a model might also consider the extent to which ratings and commentary interact, both within and across different providers. For example, is the value of commentary greater when the reviewer's evaluation is known to be extreme? Does the knowledge of a reviewer's rating bias interpretation of the commentary (or vice-versa)? As such, our research represents only one step towards a more expansive understanding of the processes by which WOM is utilized in forecasting.

In our studies, participants were allowed to elaborate on the provided WOM without any constraints on time or cognitive resources. However, such constraints are common in real-world settings, and it would be useful to consider their impact on our results. A straightforward implication of our anchoring-and-adjustment framework is that load would impede forecasts based on commentary alone to a greater extent than those based on rating (with or without commentary), due to the lack of an externally provided anchor. Future research might examine this implication directly, and address the more general issue of how cognitive constraints alter WOM-based forecasting.

In keeping with other investigations of consumer affective forecasting (Patrick et al., 2007; Wang et al., 2009), we chose to examine product categories that are more hedonic than functional in nature. We expect that the key interaction of preference similarity and WOM type would continue to operate in functional settings, although lower variance in preference similarity for functional products may restrict its impact as a moderator. Our framework also suggests that forecasts would generally improve in functional categories: on the one hand, rating-based forecasts would benefit due to the greater average preference similarity of reviewers and readers; on the other hand, commentary-based forecasts would benefit by the presence of tangible and quantifiable attributes in the commentary content, reducing errors of verbalization and simulation. However, these questions remain open, and the use of WOM for forecasting in functional categories is worthy of further investigation.

By design, the present studies provided only sparse objective information about the target products. Thus, we cannot speak to the process by which consumers may integrate more detailed product information with the rating- or commentary-based WOM that they encounter. Similarly, our studies did not include conditions in which participants received neither ratings nor commentary; therefore, we can only address the relative performance of ratings and commentary under different levels of S–R preference similarity. Finally, all four studies measured forecasting accuracy by comparing predicted and actual ratings; although this approach is common, it is subject to the concern that standards of comparison may change between forecast and consumption, reducing accuracy in a way that may not be meaningful. Tradeoff-based measures such as rankings or choices are less affected by this issue and would provide a useful complementary approach. More generally, to the extent that consumers solicit WOM under the assumption that it will ultimately improve their decisions, it would be worthwhile to test this assumption directly.

An important implication of our findings is that some forms of highly persuasive WOM may lead to undesirable outcomes for consumers. Thus a number of relevant questions present themselves: What is the relationship between the persuasiveness of WOM and its objective value as a decision aid? Do consumers learn over time to utilize WOM information more effectively, and by what process? Moreover, recent research has demonstrated contexts in which consumers consciously diverge from the choices of others, in order to assert their own uniqueness (Chan, Berger, & Van Boven, 2012; Irmak et al., 2010); however, the interplay of these contexts with WOM-based forecasting remains unexplored. From the perspective of our model, one possibility is that uniqueness motives prompt consumers to adjust further from a review-provided anchor, so that forecast error may be expected to increase (especially under high preference similarity). Each of these issues represents a promising avenue for research.

Acknowledgments

This article is based on the first author's doctoral dissertation, conducted under the supervision of the second author at the Georgia Institute of Technology. The authors would like to thank Nicholas Lurie, Koert van Ittersum, Ryan Hamilton, Jack Feldman, and members of the consumer behavior lab at Georgia Tech for their valuable feedback. In addition, the authors thank Cornelia Pechmann and three anonymous reviewers for constructive advice throughout the review process.

References

Adaval, R., & Wyer, R. S., Jr. (1998). The role of narratives in consumer information processing. *Journal of Consumer Psychology*, *7*(3), 207–245.

Archak, N., Ghose, A., & Ipeirotis, P. G. (2011). Deriving the pricing power of product features by mining consumer reviews. *Management Science*, *57*(8), 1485–1509.

Berger, J., & Schwartz, E. M. (2011). What drives immediate and ongoing word of mouth? *Journal of Marketing Research*, *48*(5), 869–880.

Berlo, D. K. (1960). *The process of communication.* San Francisco: Holt, Rinehart, and Winston.

Billeter, D., Kalra, A., & Loewenstein, G. (2011). Underpredicting learning after initial experience with a product. *The Journal of Consumer Research*, *37*(5), 723–736.

Borgida, E., & Nisbett, R. E. (1977). The differential impact of abstract vs. concrete information on decisions. *Journal of Applied Social Psychology*, *7*(3), 258–271.

Brown, J. J., & Reingen, P. H. (1987). Social ties and word-of-mouth referral behavior. *Journal of Consumer Research*, *14*(3), 350–362.

Campbell, R. S., & Pennebaker, J. W. (2003). The secret life of pronouns: Flexibility in writing style and physical health. *Psychological Science*, *14*(1), 60–65.

Chan, C., Berger, J., & Van Boven, L. (2012). Identifiable but not identical: Combining social identity and uniqueness motives in choice. *Journal of Consumer Research*, *39*(3), 561–573.

Chan, H., & Cui, S. (2011). The contrasting effects of negative word of mouth in the post-consumption stage. *Journal of Consumer Psychology*, *21*(3), 324–337.

Cheema, A., & Kaikati, A. M. (2010). The effect of need for uniqueness on word of mouth. *Journal of Marketing Research*, *47*(3), 553–563.

Chen, Y. B., Wang, Q., & Xie, J. H. (2011). Online social interactions: A natural experiment on word of mouth versus observational learning. *Journal of Marketing Research*, *48*(2), 238–254.

Chevalier, J. A., & Mayzlin, D. (2006). The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research*, *43*(3), 345–354.

Cronbach, L. (1955). Processes affecting scores on "understanding of others" and "assumed similarity". *Psychological Bulletin*, *52*(3), 177–193.

Davis, H. L., Hoch, S. J., & Ragsdale, E. K. E. (1986). An anchoring and adjustment model of spousal predictions. *Journal of Consumer Research*, *13*(1), 25–37.

De Angelis, M., Bonezzi, A., Peluso, A. M., Rucker, D. D., & Costabile, M. (2012). On braggarts and gossips: A self-enhancement account of word-of-mouth generation and transmission. *Journal of Marketing Research*, *49*(4), 551–563.

Dellarocas, C., Zhang, X., & Awad, N. F. (2007). Exploring the value of online product reviews in forecasting sales: The case of motion pictures. *Journal of Interactive Marketing*, *21*(4), 23–45.

Dickson, P. R. (1982). The impact of enriching case and statistical information on consumer judgments. *Journal of Consumer Research*, *8*(4), 398–406.

Dunning, D., Griffin, D. W., Milojkovic, J. D., & Ross, L. (1990). The overconfidence effect in social prediction. *Journal of Personality and Social Psychology*, *58*(4), 568–581.

Fieck, L., & Higie, L. A. (1992). The effects of preference heterogeneity and source characteristics on ad processing and judgments about endorsers. *Journal of Advertising*, *21*(2), 9–24.

Forman, C., Ghose, A., & Wiesenfeld, B. (2008). Examining the relationship between reviews and sales: The role of reviewer identity disclosure in electronic markets. *Information Systems Research*, *19*(3), 291–313.

Gershoff, A. D., Broniarczyk, S. M., & West, P. M. (2001). Recommendation or evaluation? Task sensitivity in information source selection. *Journal of Consumer Research*, *28*(3), 418–438.

Gershoff, A. D., Mukherjee, A., & Mukhopadhyay, A. (2003). Consumer acceptance of online agent advice: Extremity and positivity effects. *Journal of Consumer Psychology*, *13*(1–2), 161–170.

Gershoff, A. D., Mukherjee, A., & Mukhopadhyay, A. (2007). Few ways to love, but many ways to hate: Attribute ambiguity and the positivity effect in agent evaluation. *Journal of Consumer Research*, *33*(4), 499–505.

Gershoff, A. D., & West, P. M. (1998). Using a community of knowledge to build intelligent agents. *Marketing Letters*, *9*(1), 79–91.

Gigone, D., & Hastie, R. (1997). Proper analysis of the accuracy of group judgments. *Psychological Bulletin*, *121*(1), 149–167.

Gilbert, D. T., Killingsworth, M. A., Eyre, R. N., & Wilson, T. D. (2009). The surprising power of neighborly advice. *Science*, *323*(5921), 1617–1619.

Gilbert, D. T., & Wilson, T. D. (2007). Prospection: Experiencing the future. *Science*, *317*(5843), 1351–1354.

Hayes, A. F. (2012). PROCESS: A versatile computational tool for observed variable moderation, mediation, and conditional process modeling. Retrieved December 22, 2012, from: http://www.afhayes.com/public/process2012.pdf

Hoch, S. J. (1988). Who do we know: Predicting the interests and opinions of the American consumer. *Journal of Consumer Research*, *15*(3), 315–324.

Huang, P., Lurie, N. H., & Mitra, S. (2009). Searching for experience on the web: An empirical examination of consumer behavior for search and experience goods. *Journal of Marketing*, *73*(2), 55–69.

Irmak, C., Vallen, B., & Sen, S. (2010). You like what I like, but I don't like what you like: Uniqueness motivations in product preferences. *Journal of Consumer Research*, *37*(3), 443–455.

Irwin, J. R., & McClelland, G. H. (2001). Misleading heuristics and moderated multiple regression models. *Journal of Marketing Research*, *38*(1), 100–109.

Jacoby, J., Speller, D. E., & Kohn, C. A. (1974). Brand choice behavior as a function of information load. *Journal of Marketing Research*, *11*(1), 63–69.

Kahneman, D., & Snell, J. (1992). Predicting a changing taste: Do people know what they will like? *Journal of Behavioral Decision Making*, *5*(3), 187–200.

Kahneman, D., & Tversky, A. (1982). The simulation heuristic. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases*. Cambridge: Cambridge University Press.

Keller, K. L., & Staelin, R. (1987). Effects of quality and quantity of information on decision effectiveness. *Journal of Consumer Research*, *14*(2), 200–213.

Lamberton, C. P., Naylor, R. W., & Haws, K. L. (2013). Same destination, different paths: When and how does observing others' choices and reasoning alter confidence in our own choices? *Journal of Consumer Psychology*, *23*(1), 74–89.

Larrick, R. P., & Soll, J. B. (2006). Intuitions about combining opinions: Misappreciation of the averaging principle. *Management Science*, *52*(1), 111–127.

Lichtenstein, S., Slovic, P., Fischhoff, B., Layman, M., & Combs, B. (1978). Judged frequency of lethal events. *Journal of Experimental Psychology: Human Learning and Memory*, *4*(6), 551–578.

Liu, Y. (2006). Word of mouth for movies: Its dynamics and impact on box office revenue. *Journal of Marketing*, *70*(3), 74–89.

Loewenstein, G., & Adler, D. (1995). A bias in the prediction of tastes. *The Economic Journal*, *105*(431), 929–937.

Loewenstein, G., & Schkade, D. (1999). Wouldn't it be nice? Predicting future feelings. In D. Kahneman, E. Diener, & N. Schwarz (Eds.), *Well-being: The foundations of hedonic psychology* (pp. 85). New York: Russell Sage Foundation.

Mellers, B. A., Schwartz, A., Ho, K., & Ritov, I. (1997). Decision affect theory: Emotional reactions to the outcomes of risky options. *Psychological Science*, *8*(6), 423–429.

Moe, W., & Trusov, M. (2011). The value of social dynamics in online product ratings forums. *Journal of Marketing Research*, *48*(3), 444–456.

Moore, S. G. (2012). Some things are better left unsaid: How word of mouth influences the storyteller. *Journal of Consumer Research*, *38*(6), 1140–1154.

Mudambi, S. M., & Schuff, D. (2010). What makes a helpful online review? A study of customer reviews on amazon.com. *MIS Quarterly*, *34*(1), 185–200.

Muller, D., Judd, C. M., & Yzerbyt, V. Y. (2005). When moderation is mediated and mediation is moderated. *Journal of Personality and Social Psychology*, *89*(6), 852–863.

Naylor, R. W., Lamberton, C. P., & Norton, D. A. (2011). Seeing ourselves in others: Reviewer ambiguity, egocentric anchoring, and persuasion. *Journal of Marketing Research*, *48*(3), 617–631.

Newman, M. L., Pennebaker, J. W., Berry, D. S., & Richards, J. M. (2003). Lying words: Predicting deception from linguistic styles. *Personality and Social Psychology Bulletin*, *29*(5), 665–675.

Olshavsky, R. W., & Miller, J. A. (1972). Consumer expectations, product performance, and perceived product quality. *Journal of Marketing Research*, *9*(1), 19–21.

Park, D., Lee, J., & Han, I. (2007). The effect of on-line consumer reviews on consumer purchasing intention: The moderating role of involvement. *International Journal of Electronic Commerce*, *11*(4), 125–148.

Patrick, V. M., MacInnis, D. J., & Park, C. W. (2007). Not as happy as I thought I'd be? Affective misforecasting and product evaluations. *Journal of Consumer Research*, *33*(4), 479–489.

Pennebaker, J. W., Booth, R. J., & Francis, M. E. (2007). *Linguistic Inquiry and Word Count (LIWC 2007): A computer-based text analysis program.*

Pennebaker, J. W., & King, L. A. (1999). Linguistic styles: Language use as an individual difference. *Journal of Personality and Social Psychology*, *77*(6), 1296–1312.

Pennebaker, J. W., Mehl, M. R., & Niederhoffer, K. G. (2003). Psychological aspects of natural language use: Our words, our selves. *Annual Review of Psychology*, *54*(1), 547–577.

Price, L. L., Feick, L. F., & Higie, R. A. (1989). Preference heterogeneity and coorientation as determinants of perceived informational influence. *Journal of Business Research*, *19*(3), 227–242.

Read, D., & Loewenstein, G. (1995). Diversification bias: Explaining the discrepancy in variety seeking between combined and separated choices. *Journal of Experimental Psychology. Applied*, *1*(1), 34–49.

Ross, L., Greene, D., & House, P. (1977). False consensus effect — Egocentric bias in social-perception and attribution processes. *Journal of Experimental Social Psychology*, *13*(3), 279–301.

Rothwell, D. J. (2010). *In the company of others: An introduction to communication* (3rd ed.). New York: Oxford.

Ryu, G., & Han, J. K. (2009). Word-of-mouth transmission in settings with multiple opinions: The impact of other opinions on WOM likelihood and valence. *Journal of Consumer Psychology*, *19*(3), 403–415.

Schellekens, G. A. C., Verlegh, P. W. J., & Smidts, A. (2010). Language abstraction in word of mouth. *Journal of Consumer Research*, *37*(2), 207–223.

Schindler, R. M., & Bickart, B. (2012). Perceived helpfulness of online consumer reviews: The role of message content and style. *Journal of Consumer Behaviour*, *11*(3), 234–243.

Schlosser, A. E. (2011). Can including pros and cons increase the helpfulness and persuasiveness of online reviews? The interactive effects of ratings and arguments. *Journal of Consumer Psychology*, *21*(3), 226–239.

Schooler, J. W., & Engstler-Schooler, T. Y. (1990). Verbal overshadowing of visual memories: Some things are better left unsaid. *Cognitive Psychology*, *22*(1), 36–71.

Sen, S., & Lerman, D. (2007). Why are you telling me this? An examination into negative consumer reviews on the web. *Journal of Interactive Marketing*, *21*(4), 76–94.

Sengupta, J., & Fitzsimons, G. J. (2000). The effects of analyzing reasons for brand preferences: Disruption or reinforcement? *Journal of Marketing Research*, *37*(3), 318–330.

Shafir, E., Simonson, I., & Tversky, A. (1993). Reason-based choice. *Cognition*, *49*(1), 11–36.

Shiv, B., & Huber, J. (2000). The impact of anticipating satisfaction on consumer choice. *Journal of Consumer Research*, *27*(2), 202–216.

Simonson, I. (1990). The effect of purchase quantity and timing on variety-seeking behavior. *Journal of Marketing Research*, *27*(2), 150–162.

Solomon, M. R. (1986). The missing link — Surrogate consumers in the marketing chain. *Journal of Marketing*, *50*(4), 208–218.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124–1131.

Wang, J., Novemsky, N., & Dhar, R. (2009). Anticipating adaptation to products. *Journal of Consumer Research*, *36*(2), 149–159.

Watts, D. J., & Dodds, P. S. (2007). Influentials, networks, and public opinion formation. *Journal of Consumer Research*, *34*(4), 441–458.

Weiss, A. M., Lurie, N. H., & MacInnis, D. J. (2008). Listening to strangers: Whose responses are valuable, how valuable are they, and why? *Journal of Marketing Research*, *45*(4), 425–436.

Wilson, T. D., & Gilbert, D. T. (2003). Affective forecasting. *Advances in Experimental Social Psychology*, *35*(35), 345–411.

Wilson, T. D., & Schooler, J. W. (1991). Thinking too much: Introspection can reduce the quality of preferences and decisions. *Journal of Personality and Social Psychology*, *60*(2), 181–192.

Wilson, T. D., Wheatley, T., Meyers, J. M., Gilbert, D. T., & Axsom, D. (2000). Focalism: A source of durability bias in affective forecasting. *Journal of Personality and Social Psychology*, *78*(5), 821–836.

Wood, S. L., & Bettman, J. R. (2007). Predicting happiness: How normative feeling rules influence (and even reverse) durability bias. *Journal of Consumer Psychology*, *17*(3), 188–201.

Zhao, M., Hoeffler, S., & Dahl, D. W. (2009). The role of imagination-focused visualization on new product evaluation. *Journal of Marketing Research*, *46*(1), 46–55.

Zhao, M., & Xie, J. H. (2011). Effects of social and temporal distance on consumers' responses to peer recommendations. *Journal of Marketing Research*, *48*(3), 486–496.

Zhu, F., & Zhang, X. Q. (2010). Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics. *Journal of Marketing*, *74*(2), 133–148.