CONSUMER JUDGMENT AND FORECASTING USING ONLINE WORD-OF-MOUTH

A Dissertation
Presented to
The Academic Faculty

by

Stephen Xihao He

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
Scheller College of Business

Georgia Institute of Technology
August 2012

CONSUMER JUDGMENT AND FORECASTING USING ONLINE WORD-OF-MOUTH

Approved by:

Dr. Samuel D. Bond, Advisor
College of Management
Georgia Institute of Technology

Dr. Koert van Ittersum
College of Management
Georgia Institute of Technology

Dr. Nicholas H. Lurie
School of Business
University of Connecticut

Dr. Jack M. Feldman
School of Psychology
Georgia Institute of Technology

Dr. Ryan Hamilton
Goizueta Business School
Emory University

Date Approved: June 6, 2012

To Mom, Dad, and Wen, for their love and support

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to Samuel Bond for his guidance, support, and advice throughout my doctoral program. He not only taught me the craft of how to conduct and write experimental research, but also gave me the freedom to explore the problem from new perspectives. The research projects documented here would not have been possible without the invaluable contribution of his time and effort.

This dissertation also owes a great deal to other members in my committee, Koert van Ittersum, Nicholas Lurie, Jack Feldman, and Ryan Hamilton, who provided great encouragement and much constructive advice in the development of ideas, study designs, and theory building.

I am also grateful to the marketing faculty and doctoral students who showed interest in my work and helped me along the way. Special thanks goes to Ajay Kohli, Jun Kim, Xiaoqin Jing, and Goutam Challagalla, who offered great feedbacks when these ideas were in their infancy, as well as Na Wen, Dezhi Yin, Zoey Chen, Daniel Sheehan, and Mike Hair, who helped out my data collection process considerably.

Finally, I would like to express appreciation to my parents, Yegang He and Yaping Zhou, and my fiancée, Wen Wen, for their love, understanding, patience, and encouragement throughout this endeavor. I dedicate this dissertation to them.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

Empowered by information technology, modern consumers increasingly rely upon online word-of-mouth (WOM—e.g., product reviews) to guide their purchase decisions. This dissertation investigates how WOM information is processed by consumers and its downstream consequences.

The quality of consumer decisions often depends on accurate forecasting of future enjoyment. Chapter 2 of the dissertation explores the value of specific types of word-of-mouth information (e.g., numeric ratings, text commentary, or both) for making forecast. After proposing an anchoring-and-adjustment framework for the utilization of WOM to inform consumer forecasts, I support this framework with a series of experiments. Results demonstrate that the relative forecasting advantage of different information types is a function of the extent to which consumer and reviewer have similar product-level preferences (‘source-receiver similarity’).

Consumers are more likely than ever to encounter a mixture of positive and negative WOM. Chapter 3 of the dissertation investigates the process by which dispersion—the degree to which opinions are divided for a product or service—in WOM is interpreted. Using an attribution-based approach, I argue that the effect of WOM dispersion is dependent on the perceived cause of that dispersion, which is systematically related to perceptions of preference heterogeneity in a product category. For products for which preferences are expected to vary, dispersion is likely to be attributed to the reviewers rather than the product itself, and therefore tolerated. I provide evidence for my

proposal in a series of experiments where WOM dispersion is manipulated and respondents make choices and indicate purchase intentions.

CHAPTER 1

INTRODUCTION

For the vast majority of consumer decisions, others have already experienced one or more of the options under consideration and shared their own opinions through online word-of-mouth (WOM—e.g., product reviews). Empowered by information technology, modern consumers increasingly rely upon such information to guide their purchase decisions. The topic has gained attention from marketing researchers interested in its effects on decision making (Cheema and Kaikati 2010; Irmak, Vallen, and Sen 2010; Naylor, Lamberton, and Norton 2011; Weiss, Lurie, and MacInnis 2008), sales (Chevalier and Mayzlin 2006; Chintagunta et al. 2010; Clemons et al. 2006; Dellarocas et al. 2007; Li and Hitt 2008; Moe and Trusov 2011; Moon et al. 2010; Sun 2011; Zhu and Zhang 2010), and related variables. I am interested in how WOM information is processed by consumers and its downstream consequences, and this dissertation contributes to the understanding of these important questions by investigating the forecasting value of WOM and the risks embedded in WOM dispersion.

One objective criterion for evaluating the usefulness of WOM to consumers is the extent to which it enables accurate forecasting of future enjoyment. Based on this criterion, Chapter 2 of the dissertation explores the value of specific types of WOM information (numeric ratings, text commentary, or both) for making forecasts. I propose an anchoring-and-adjustment framework in which the relative value of different information types is a function of the extent to which consumer and reviewer have similar product-level preferences (‘source-receiver similarity’). In particular, I argue that numeric ratings are processed by a heuristic of 'assumed similarity' that results in limited adjustment, while text commentary invokes a process of mental simulation that produces greater (but not necessarily better) adjustment. This framework allows me to derive conditions under which ratings, commentary, or their combinations are more useful for prediction. A series of experiments demonstrates that the forecasting advantage of ratings is restricted to cases where source-receiver similarity is high (e.g., products with relatively homogenous preferences). Most importantly, as source-receiver similarity decreases, the usefulness of ratings declines substantially, but the usefulness of commentary remains intact. As predicted by my framework, the data also reveal conditions where the presence of ratings and commentary together actually inhibit forecasts. Finally, despite the presumed benefits of aggregating WOM through use of 'average' ratings, I show that when preferences are heterogeneous, forecasts based on averages of ratings may underperform those based on a single review, especially when commentary is provided.

Regardless of the specific format of WOM involved, the advance of social and mobile technologies has ensured that online consumers inevitably encounter a mixture of positive and negative reviews. However, the magnitude of this disagreement can vary dramatically across and within product categories. Both intuition and existing evidence suggest that because uncertainty is usually undesirable, consumers will tend to favor products with consistent WOM. Chapter 3 of the dissertation presents a more nuanced perspective. In particular, I investigate WOM using an attribution-based approach and provide experimental evidence for the process by which divided opinions are interpreted. I argue that WOM dispersion signals (in)consistency across consumption incidences, but whether this (in)consistency is attributed to the product or the reviewers will depend on the decision context (e.g., product type, positioning, etc.). A reviewer-oriented attribution draws attention away from the focal product while emphasizing the ability of the consumer to control their outcome. Consequently, consumers will be more tolerant of WOM dispersion when they attribute the dispersion to reviewer idiosyncrasies. I provide evidence for this approach in a series of experiments where respondents provide choices and intentions in response to decisions with varying levels of WOM dispersion. These experiments demonstrate that when products are characterized by heterogeneous preferences or when reviewers are known to differ from one another, dispersion is less detrimental to choice likelihood and purchase intention. Thus, they provide important insights into consumer WOM, risk perception, and the causal attribution of uncertainty.

CHAPTER 2

WORD-OF-MOUTH AND THE FORECASTING OF CONSUMPTION ENJOYMENT

2.1 Introduction

"Enjoying the joys of others and suffering with them—these are the best guides for man."

-Albert Einstein

Satisfactory purchase decisions often depend on the ability to accurately forecast future consumption experience. Ideally, online shopping environments facilitate the forecasting process by increasing access to informative word-of-mouth (WOM), through which product information is transmitted between consumers (Brown and 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 or less useful for forecasting purposes. This essay addresses these issues.

I focus on two common forms of WOM: summary ratings and review commentary (i.e., text reviews). Ratings and commentary represent a class of 'surrogate' WOM in which the usage experience and opinions of peer consumers are presented directly. An emerging research stream has documented the influence of product ratings on sales (Chevalier and Mayzlin 2006; Liu 2006; Moe and Trusov 2011), and a separate literature has investigated the economic impact of commentary (Archak, Ghose, and

Ipeirotis 2011; Park, Lee, and Han 2007), but there has been little research directly comparing these types of information on consumer outcomes. In contrast to prior work examining WOM from the viewpoint of firms or retailers, my research explicitly adopts a consumer perspective, focusing on the utilization of WOM to predict future enjoyment and satisfaction. Intuitively, marketers and consumers might expect a numeric 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 and Wyer 1998). However, research in affective forecasting reveals a litany of simulation errors which cast doubt on this assumption: in fact, a single peer rating is sometimes more helpful for prediction than detailed descriptions of an experience (Gilbert, Killingsworth, Eyre, and Wilson 2009).

My contribution is fourfold. First, I amend the affective forecasting literature by examining how consumers utilize word-of-mouth from others to inform their forecasts. To do so, I 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 me to examine the relative utility of ratings, commentary, or their combination for prediction. Interestingly, and counter to the notion that "more information is better," I demonstrate that rating and commentary together are sometimes less useful than one of these alone. Third, I incorporate a highly relevant variable, heterogeneity in preferences, and demonstrate that heterogeneity affects the value of WOM in a manner congruent with my framework. Finally, given that consumers often utilize 'average' ratings as a decision aid, I consider the implications of this aid for forecast accuracy.

2.2 Conceptual Background

Forecasting Future Experience

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 and actual experience. The topic has received substantial scholarly attention (Hoch 1988; Loewenstein and Adler 1995; Patrick, MacInnis, and Park 2007; Wang, Novemsky, and Dhar 2009). A robust finding of this work is that individuals are poor at making affective forecasts, especially for hedonic experiences (Billeter, Kalra, and Loewenstein 2011; Kahneman and Snell 1992; Read and Loewenstein 1995; Simonson 1990; Wilson, Wheatley, Meyers, Gilbert, and Axsom 2000; Wood and Bettman 2007). Most commonly, forecasting errors are attributed to faulty simulation of future experiences (Gilbert and Wilson 2007; Zhao, Hoeffler, and Dahl 2009), and prescriptive advice often aims at improving the simulation process.

Online WOM has gained increasing attention from consumer researchers interested in its effects on decision making (Weiss, Lurie, and MacInnis 2008), purchase intention (Chevalier and Mayzlin 2006), and related variables. In keeping with recent research on the use of peer knowledge in personal predictions (Gilbert, et al. 2009), this paper focuses on WOM as a means of improving consumer forecasts. In particular, I focus on instances in which consumers relay their own, experience-based evaluation of goods and services (e.g., consumer reviews available through vendors or third parties). Although such WOM offers clear informational benefits to consumers, its availability also adds complexity to the decision environment and increases consumers' processing burden (Ansari and Mela 2003).

Word-of-Mouth: Ratings and Commentaries

Scholars have conceptualized the communication process using a variety of frameworks but most involve four fundamental elements: source, receiver, message, and channel (Berlo 1960, Rothwell 2010). In terms of these elements, I focus in particular on two issues: 1) the message format, and 2) the relationship between source and receiver. Regarding format, ratings and commentaries represent two common implementations of WOM that have been widely studied (though rarely from the consumer's perspective—Chevalier and Mayzlin 2006; Dellarocas, Zhang, and Awad 2007; c.f. Park, et al. 2007). Almost all WOM-enabled retailers provide some form of peer-generated product or service evaluation, usually in the form of an 'overall' rating scale intended to summarize positive and negative aspects with a single number. Although consumers may disagree on the perceptual meaning of specific ratings, they at least know the range of possible values and recognize that larger values connote more positive evaluations. Under ideal conditions, an overall rating conveys the reviewer's opinion accurately, with minimal effort required from the reader.

In contrast to an overall rating, a textual commentary provides a far richer context and often includes vivid and concrete content that enhances mental simulation (Adaval and Wyer 1998; Dickson 1982). Although the helpfulness of commentary varies by its depth and readability (Archak, et al. 2011; Mudambi and Schuff 2010), it typically contains both objective and subjective content relevant to the decision. Moreover, commentary often provides readers with a variety of reasons underlying the author's evaluation; these reasons may in turn be utilized by the reader to resolve decision conflict or to justify their own choices (Shafir, Simonson, and Tversky 1993). Hence, it is natural for consumers to assume that commentary will be more useful than a simple overall rating,[1] and evidence indicates that consumers provided with both ratings and written content tend to rely heavily on the latter (Freedman 2008; Schlosser 2011).

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 may overemphasize certain aspects that are easier to recall or verbalize (Schooler and Engstler-Schooler 1990). In addition, it may contain reasoning that is ad hoc or inconsistent with the reviewer's attitude (Sengupta and Fitzsimons 2000; Wilson and Schooler 1991). In contrast, ratings are concise and easily understood, representing the evaluations of diverse peers on a common scale (though it need not true that they utilize the scale identically). Under ideal conditions, an overall rating conveys the reviewer's opinion accurately, with minimal effort required from the reader. In one prominent example (Gilbert et al. 2009), even a single peer's evaluation of an experience

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

was more useful to forecasters than an objective description. In sum, it is unclear whether rating or commentary is inherently 'superior' from an informational standpoint. Therefore, I focus on factors affecting their relative value.

Source-Receiver Preference Similarity and Forecast Accuracy

Substantial evidence indicates that consumers look for – and are persuaded by – information provided by similar peers (Forman, Ghose, and Wiesenfeld 2008; Gershoff, Mukherjee, and Mukhopadhyay 2007; Price, Feick, and 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, I propose a crucial moderator: *source-receiver preference similarity*. I define source-receiver preference similarity as the overlap in product-specific preferences of a reviewer and a prospective consumer. In principle, this construct captures the difference in utility function between the two (i.e., their weighting and valuation of product attributes).

Ideally, source-receiver preference similarity could be directly measured as the actual difference in product evaluations between source and receiver, and I utilize this approach in two of my studies. In the marketplace, however, actual product evaluations of prospective consumers cannot be known in advance at the individual level. 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 good vary within a population (Gershoff and West 1998; Price, et al. 1989). For some product categories, consumers are likely to have reasonable lay theories of preference heterogeneity (e.g., preferences will vary less widely for functional goods than taste goods); in other cases, heterogeneity may be inferred from information provided by the vendor (e.g., a graph of ratings dispersion) or a third party (e.g., Consumer Reports).

Applied to the current context, preference heterogeneity provides a useful proxy for source-receiver similarity. For products characterized by high preference heterogeneity, any one reviewer's preferences are likely to differ not only from those of other reviewers, but also from those of prospective consumers; i.e., average levels of source-receiver similarity will be low. Therefore, a prospective consumer is unlikely to be 'matched' with a reviewer whose preferences are similar. For products characterized by homogeneous preferences, evaluations differ little across consumers; i.e., average levels of source-receiver similarity are high. Therefore, a prospective consumer is likely to be 'matched' with a reviewer whose preferences are similar. My first study examines the impact of preference heterogeneity on WOM-based predictions.

In developing predictions for when ratings will outperform commentary, and vice-versa, I c adopt an anchoring-and-adjustment framework (Lichtenstein, Slovic, Fischhoff, Layman, and Combs 1978; Tversky and Kahneman 1974). In my framework, receivers estimate the source's (reviewer's) evaluation, then adjust that evaluation based on the extent to which they believe their own preferences align with those of the reviewer (c.f. egocentric models for prediction of others' preferences – Davis et al. 1986). When WOM consists merely of an overall rating, that rating serves as a natural and readily available forecasting anchor; indeed, consumers often rely on others' ratings to estimate their own (Irmak, Vallen, and Sen 2010). Note that even when preference similarity with the reviewer is unknown, consumers will often adjust their own prediction from that of the reviewer. For example, extremity aversion may provoke an adjustment towards neutrality; general optimism or pessimism may provoke adjustment upward or downward; experience with the product category may provoke adjustment consistent with that experience. Typically, however, the extent of any adjustment will be limited, a phenomenon which Cronbach (1955) labeled "assumed similarity." Moreover, Naylor et al (2011) have shown that consumers often perceive themselves as highly similar to an ambiguous information source, whether or not such perceptions are warranted. As a result, it is expected that: a) minimal adjustment will occur, and b) any adjustment that does occur will be of limited value. Therefore, when forecasts are based on a rating alone, error will be minimized when source-receiver preference similarity is high and maximized when source-receiver preference similarity is low. Consequently, rating-based forecasts should be more (less) accurate than commentary for product categories in which preferences are more homogeneous (heterogeneous).

In contrast, I expect forecasts based on commentary to be less dependent on similarity between writer and reader. Although commentary lacks a direct indicator of the reviewer's evaluation, it provides descriptive semantic content conducive to visualization and mental simulation (Gilbert and Wilson 2007; Kahneman and Tversky 1982). In written 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 thus contrast the reviewer's preferences with their own. Assume, for example, that a reviewer speaks favorably of an apartment complex but laments its pool quality. A prospective renter who does not swim may use this information 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 upward. Because source-receiver similarity is identified and adjusted for, its impact is reduced.

Figure 1 illustrates the process that I propose. In the first panel (A1-A2), consumers exposed to mere rating information use that information as an anchor for their forecasts, and make limited adjustment. In the second panel (B1-B2), consumers who read commentary make an estimate of the reviewer's rating as their anchor, and then use similarity cues in the commentary to adjust that anchor. In the third panel (C1-C2), where both rating and commentary are available, consumers anchor on the rating and utilize the commentary to make similarity-based adjustment. The figure illustrates that when a reviewer and reader have similar preferences, commentary will tend to lead to more forecast error due to noise in the estimation / adjustment processes. In contrast, when a reviewer and reader have dissimilar preferences, the rating results in a misleading anchor, while commentary enables readers to make adjustments based on awareness that their criteria are different than those of the reviewer.

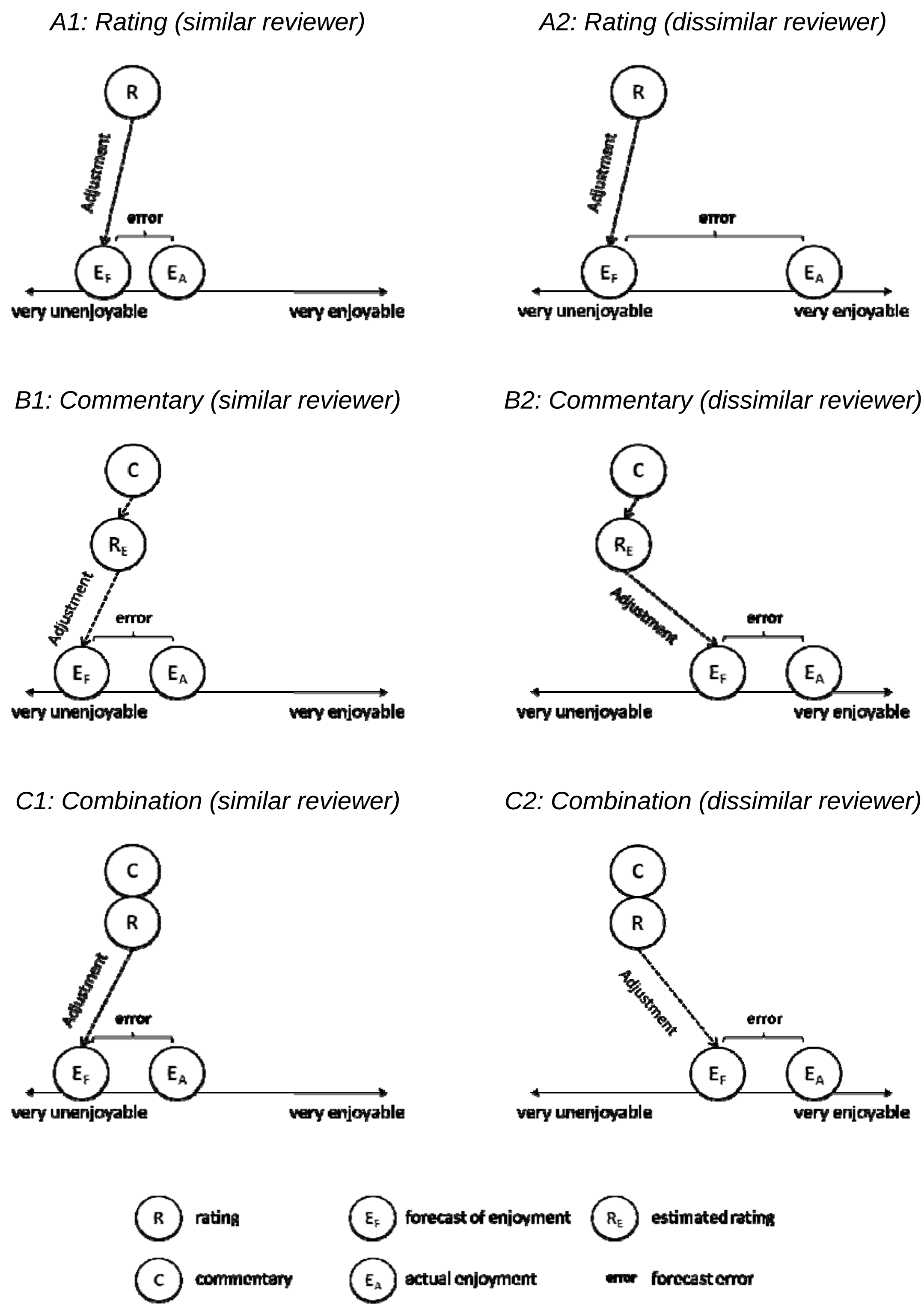


Figure 1: Graphic Interpretation of the Forecasting Process

In sum, my model predicts that the effect of preference heterogeneity on WOM-based forecasts will depend on the type of WOM involved. Being 'matched' with a similar reviewer is critical in the case of ratings, but has little impact on the value of commentary; thus, the advantages of commentary will be greatest for products with heterogeneous preferences. Note that this interaction model does not predict a general superiority of commentary over ratings. Accounting for interpersonal differences is inherently difficult (Davis, Hoch, and Ragsdale 1986; Hoch 1988), and estimates based on mental simulation are subject to misinterpretation, egocentric bias, focalism, etc. (Wilson, et al. 2000). When readers and reviewers tend to have different preferences, these errors in commentary processing will be negligible compared to its benefits for adjustment, and forecasts based on commentary will outperform those based on ratings. However, when reviewers and readers tend to share similar preferences, errors in commentary processing will remain, but the 'natural anchor' of a rating will be very useful for prediction. In sum, I predict the following:

H1: Commentary information leads to more accurate forecasts than rating information when preference heterogeneity is high, but this advantage diminishes as preference heterogeneity decreases.

Many review forums provide rating and commentary information together. In these cases, consumers receive not only an "error-free" anchor for the reviewer's evaluation, but also evidence of underlying preferences that can be used to infer necessary adjustment. Although intuition suggests a synergy by which the combination is more useful than either rating or commentary alone, my framework implies that this need not obtain. Based on abundant evidence that individuals tend to overweight vivid or case-based information compared to statistical or numeric information (Borgida and Nisbett 1977; Dickson 1982; Schlosser 2011), I expect that processing in the combined condition will focus heavily on the commentary portion. Hence, forecasts based on combined information will still be subject to the errors of interpretation and simulation described above. When readers and reviewers tend to have different preferences, these errors are trivial compared to the benefits of commentary for adjustment. However, as reviewers and readers become more similar, the need for adjustment diminishes, while the cost of these errors remains; in cases of extreme homogeneity, forecasts based on rating and commentary together may even underperform those based on a rating alone:

H2: Combined information leads to more accurate forecasts than rating information when preference heterogeneity is high, but this advantage diminishes as preference heterogeneity decreases.

The relative value of combined information versus commentary alone is unclear. Although the addition of an overall rating to a reviewer's commentary ensures an accurate anchor, the rating may also introduce new error in the interpretation/adjustment process. For example, knowledge of a reviewer's overall evaluation may discourage attempts by the reader to process commentary carefully, or may create a halo effect which biases its interpretation. Therefore, I make no formal prediction regarding the relative value of combined information versus commentary alone. Rather, I expect that the two scenarios will tend to invoke similar processing patterns and result in similar levels of performance. Study 1 offers an empirical test of this prediction.

My discussion thus far has been restricted to WOM from a single reviewer. However, aggregate 'average' ratings are often provided by online vendors, and it is reasonable to expect that average ratings would be especially valuable for forecasting. This intuition is consistent with the notion of the "wisdom of crowds," by which aggregated group judgments are often more accurate than those of individuals (Gigone and Hastie 1997; Larrick and Soll 2006). However, unlike the objective judgments shown to benefit from aggregation, product preferences are inherently idiosyncratic. An average rating is akin to a 'preference of crowds,' whose usefulness for forecasting depends on the dispersion of those preferences. When preference heterogeneity is high, an average rating is likely to be a poor anchor for predicting one's own evaluation, but when preference heterogeneity is low, the anchor is likely to be more accurate. In contrast – and as discussed above – the benefits and costs of commentary are largely independent of preference heterogeneity. Thus, I predict the following:

H3: a. Commentary information leads to more accurate forecasts than average ratings when preference heterogeneity is high, but this advantage diminishes as heterogeneity decreases.

b. Combined information leads to more accurate forecasts than average ratings when preference heterogeneity is high, but this advantage diminishes as heterogeneity decreases.

2.3 Overview of Studies

All four of my studies utilize a matched-pair paradigm (Gilbert, et al. 2009), in which participant ‘receivers’ (readers) are assigned randomly to ‘sources’ (reviewers) from a preliminary session. Each of the studies has three distinct components: 1) collection of WOM from preliminary reviewers who undergo the consumption experience, 2) construction of forecasts by readers who receive that WOM, and 3) actual evaluations of the consumption experience by the same readers. My key dependent variable is forecast error, defined as the absolute discrepancy between forecasts and evaluations. Given that hedonic products are typically hard to quantify, difficult to describe, and associated with low forecasting accuracy (Huang, Lurie, and Mitra 2009; Patrick, et al. 2007; Wang, et al. 2009), all four studies utilize hedonic stimuli (music and jellybeans). To ensure that participants rely solely on WOM, only sparse product information is presented (Gershoff, Broniarczyk, and West 2001). Key independent variables include type of WOM, source-receiver preference similarity (measured or manipulated), and product-level preference heterogeneity.

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

2.4 Study 1

Study 1 examined the influence of different types of WOM on forecast accuracy at different levels of preference heterogeneity Participants in the study were asked to predict their enjoyment of different jellybeans based on provided WOM. 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 heterogeneity; 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, $F(22, 22) = 1.93$, $p = .07$). Two other flavors were chosen as fillers (root beer: M = 55.83, SD = 24.80; pear: M = 48.96, SD = 29.45), in order 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 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 it's 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 marshmellow 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 a piece of WOM collected during the pretest, then forecast how much they would enjoy the jellybean on the 100-pt enjoyment scale. The study constituted a 2 (preference heterogeneity: *low* vs. *high*) x 4 (WOM type: *rating* vs. *commentary* vs. *combination* vs. *average rating*) mixed design. As described above, preference heterogeneity was manipulated within-subjects by use of two flavors (cinnamon and vanilla). WOM type was manipulated between-subjects following a random-pairing approach common in social prediction research (Dunning, Griffin, Milojkovic, and Ross 1990; Gilbert et al 2009): In the *rating* condition, each participant viewed one rating, randomly chosen, from those collected earlier; 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 *average rating* condition, each participant viewed the average rating of the pretest group. With the exception of the *avg. rating* condition, the WOM given to a participant for each jellybean was provided by a different reviewer, and randomization was constrained to ensure that ratings and commentaries from each reviewer were presented equally often. In addition to making their forecasts, participants answered two process-related questions (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 the same (100-pt) enjoyment scale.

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

Perceived reviewer enjoyment / Adjustment. For each jellybean, participants were asked to estimate how much the reviewer enjoyed the jellybean on the 100-pt scale. In the *rating*, *combination*, and *avg. rating* conditions, this measure verified that the rating was encoded accurately, while in the *commentary* condition, it captured participants' perceptions 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 chose to deviate from the opinion provided by the reviewer.

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").

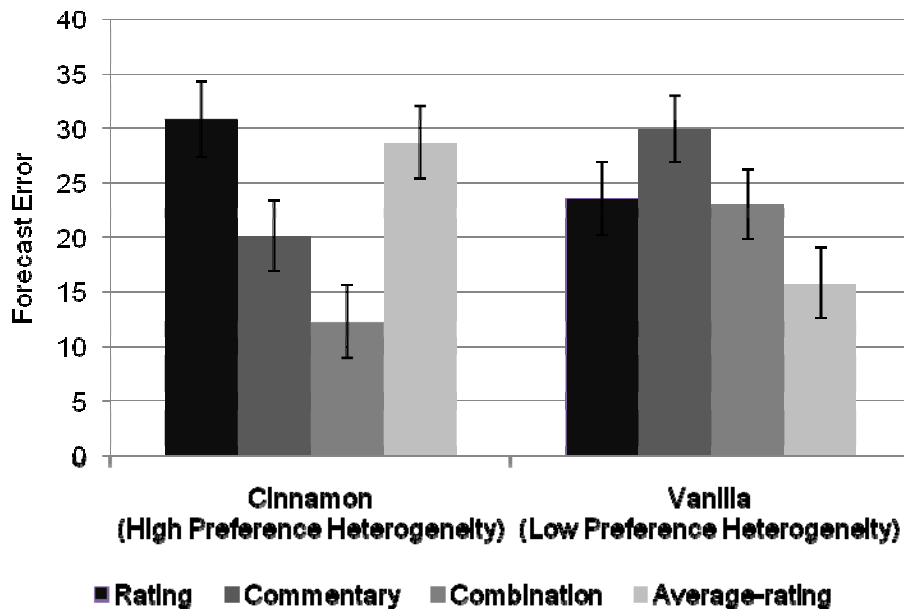
Results and Discussion

Prior to the main analysis, I first compared the actual source-receiver preference similarity observed for the high-heterogeneity stimulus (cinnamon) and low-heterogeneity stimulus (vanilla). Preference similarity was computed by subtracting from

100 the absolute difference between the evaluation of each participant and the evaluation of his/her review author (so that bigger numbers indicate greater similarity). As expected, actual source-receiver similarity was significantly less for the high-heterogeneity stimulus than the low-heterogeneity stimulus: $M = 67.16$ vs. 75.14, $F(1, 228) = 7.94$, $p < .01$).

Mean forecast errors are plotted in Figure 2, and Table 2 summarizes forecast error for all four studies in the paper. H1 and H2 were tested by using a mixed-effects model to predict forecast error as a function of WOM type, jellybean flavor, and their interaction. Analyses revealed a main effect for WOM type ($F(3, 228) = 3.14$, $p = .03$); more importantly, and consistent with predictions, analyses also revealed a significant interaction between WOM type and heterogeneity ($F(3, 228) = 6.77$, $p < .001$), 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. *average rating* $F(1\ 228) = 12.48$, $p < .001$; *combination* vs. *average rating* $F(1\ 228) = 13.13$, $p < .001$).

Follow-up comparisons revealed a pattern consistent with H1 and H2. For the flavor having high preference heterogeneity (cinnamon), forecast error was higher 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 ($M = 12.37$; $F(1, 228) = 14.94$, $p < .001$). However, for the flavor having low preference heterogeneity (vanilla), 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 ($M = 23.10$; $F(1, 228) <1$). In sum, commentary was clearly more useful than ratings when preferences were diverse, but this advantage was not observed when preferences were more homogeneous.



Notes: Forecast error measures the absolute value of the difference between predictions and subsequent evaluations.

Figure 2: Study 1: Forecast Error by WOM Type and Preference Heterogeneity

Table 2: Study 1: Forecast Error by WOM Type and Experimental Condition

Study	Condition	Rating	Commentary	Combination	Avg. rating
1	Cinnamon (high heterogeneity)	30.93 (3.48)	20.22 (3.20)	12.37 (3.31)	28.76 (3.36)
	Vanilla (low heterogeneity)	23.63 (3.34)	30.00 (3.07)	23.10 (3.17)	15.83 (3.22)
2a	Similarity (25th percentile)	27.64 (1.48)	17.95 (1.34)	21.83 (1.27)	n/a
	Similarity (50th percentile)	18.09 (1.15)	16.39 (1.15)	17.88 (1.18)	n/a
	Similarity (75th percentile)	10.80 (1.44)	15.20 (1.51)	14.86 (1.52)	n/a
2b	Low similarity	25.63 (2.04)	22.31 (1.98)	19.95 (1.82)	n/a
	High similarity	11.48 (1.67)	17.58 (1.86)	16.63 (2.09)	n/a
3	Informed	20.68 (2.51)	24.44 (2.30)	22.33 (2.32)	n/a
	Uninformed	28.24 (2.29)	25.03 (2.35)	23.13 (2.35)	n/a
	Misinformed	34.90 (2.37)	24.77 (2.51)	24.86 (2.67)	n/a

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

According to the framework presented earlier, consumers given commentary alone and those given combined information should be similarly impacted by preference heterogeneity, as both groups will utilize the written content to infer similarity. In line with this argument, I observed no partial interaction between flavor and WOM type for *commentary* and *combination* conditions ($F(1, 228) < 1$). This result was replicated in the next two studies and is clearly evident in Table 2: although the relative performance of *commentary* and *combination* conditions varied across studies, they responded similarly to underlying variance in preferences. While tentative, these findings support my view that participants given commentary (with or without a rating) integrate source-receiver similarity into their forecasts in a comparable manner.

To test H3a and H3b, I compared forecasts based on WOM from a single reviewer to those based on an average rating. For cinnamon, forecast error in the *average rating* condition ($M = 28.76$) was marginally higher than that in the single reviewer *commentary* condition ($M = 20.22$; $F(1, 228) = 3.39$, $p = .07$) and significantly higher than that in the single reviewer *combination* condition ($M = 12.37$; $F(1, 228) = 12.09$, $p = .001$). In other words, when preferences were diverse, an average rating was no more helpful (and was often less helpful) than the review of a single peer. However, when preferences were more homogeneous, this pattern reversed: for vanilla, forecast error in the *average rating* condition ($M = 15.83$) was significantly lower than that in the *commentary* condition ($M = 30.00$; $F(1, 228) = 10.14$, $p < .01$) and directionally lower than that in the *combination* condition ($M = 23.10$; $F(1, 228) = 2.59$, $p = .11$).

Mean adjustment by condition is summarized in Table 3, for study 1 along with the other three studies. My framework states that when given rating information alone, consumers will have little reason to deviate from that rating (i.e., adjustment will be minimal), but when given commentary information, participants will use that commentary to infer potential differences with the reviewer. A mixed-effect model was used to predict adjustment from WOM type, jellybean flavor, and their interactions. Analysis identified main effects for WOM type ($F(3, 228) = 8.02$, $p < .001$) and flavor ($F(1, 228) = 8.06$, $p < .01$). 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 ($M = 20.98$, $F(1, 228) = 4.67$, $p = .03$; $M = 20.35$, $F(1, 228) = 3.75$, $p = .05$). In addition, adjustment in the *avg. rating* condition was even lower than that in the *rating* condition ($M = 7.31$; $F(1, 228) = 4.14$, $p = .04$), indicating that participants were even more likely to conform to aggregate opinions (Watts and Dodds 2007).

Table 3: Study 1-3: Adjustment and Correlation Between Adjustment and Actual Similarity, by WOM Type

Study	WOM type	*n*	Adjustment		Correlation	
			M	*SD*	*r*	*sig.*
1	Rating	54	14.02	2.37	-0.07	0.62
	Commentary	64	20.98	2.18	-0.28	0.03
	Combination	60	20.35	2.25	-0.48	0.00
	Avg. rating	58	7.31	2.29	-0.01	0.95
2a	Rating	158	11.87	1.29	-0.17	0.04
	Commentary	162	23.98	1.25	-0.18	0.02
	Combination	157	23.17	1.28	-0.41	0.00
2b	Rating	132	13.96	1.29	-0.14	0.11
	Commentary	120	19.01	1.33	-0.19	0.04
	Combination	117	16.71	1.36	-0.37	0.00
3	Rating-informed	68	15.92	2.62	-0.30	0.01
	Rating-uninformed	80	12.58	2.38	-0.18	0.10
	Rating-misinformed	76	20.52	2.45	0.39	0.01
	Commentary-informed	80	27.06	2.38	-0.28	0.01
	Commentary-uninformed	76	26.82	2.44	-0.07	0.55
	Commentary-misinformed	68	25.19	2.61	-0.12	0.33
	Combination-informed	79	24.06	2.42	-0.47	0.00
	Combination-uninformed	76	19.14	2.44	-0.18	0.11
	Combination-misinformed	61	24.38	2.75	-0.22	0.09

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

Under my framework: the ability to form similarity inferences from commentary content and adjust one's forecast accordingly is most helpful in situations where source and receiver have different preferences. Table 3 summarizes correlations between adjustment and actual similarity. In conditions where commentary was available (*commentary* and *combination*), adjustment was negatively correlated with actual similarity (r = -.48 and r = -.28), as would be expected if readers were able to infer preference similarity from the commentary. Presumably, it is this ability to infer and adjust that explains the negligible impact of similarity on forecast accuracy among these participants. However, no correlation between adjustment and similarity observed in the *rating* and *avg. rating* conditions (r = -.07 and r = -.01, NS). Consistent with the results above, these findings strongly support my contention that forecasts based on WOM are influenced by perceptions of source-receiver preference similarity, even when those perceptions are inaccurate.

Table 4 presents forecast confidence for study 1 along with the other three studies, and the table also summarizes correlations between forecast confidence and error. Responses to the confidence measure revealed a main effect of WOM type (F(3, 228) = 15.65, p < .001). Follow-up comparisons revealed 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 both 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). These findings are consistent with the arguments presented earlier that consumers will believe commentary to be useful, based on its vividness, detail, provision of reasons, etc.

To the extent that participants were able to gauge the usefulness of WOM, confidence in forecasts would be expected to show a strong negative correlation with actual forecast error, particularly in the *commentary* and *combination* conditions. However, the overall correlation between confidence and forecast error was small (r = -.16), as was the correlation in each condition. This general pattern was replicated in studies 2a, 2b, and 3: confidence-error correlations were consistently near zero, and in some cases even positive (e.g., the *commentary* condition of study 2b). Together, these findings call into question consumers' ability to recognize the usefulness of the WOM they are provided.

Table 4: Study 1-3: Forecast Confidence and Correlation Between Forecast Confidence and Forecast Error, by WOM Type

Study	WOM type	*n*	Forecast confidence		Correlation	
			M	*SD*	*r*	*sig.*
1	Rating	54	3.93	0.19	-0.35	0.01
	Commentary	64	5.52	0.17	-0.14	0.26
	Combination	60	5.18	0.18	-0.10	0.43
	Avg. rating	58	4.45	0.18	0.04	0.74
2a	Rating	158	3.52	0.12	0.00	0.99
	Commentary	162	4.63	0.12	0.04	0.61
	Combination	157	4.79	0.12	-0.04	0.64
2b	Rating	132	2.78	0.14	0.11	0.21
	Commentary	120	4.97	0.14	0.24	0.01
	Combination	117	4.71	0.14	-0.18	0.05
3	Rating-informed	68	4.37	0.17	-0.04	0.75
	Rating-uninformed	80	3.86	0.16	0.13	0.24
	Rating-misinformed	76	4.29	0.16	0.19	0.09
	Commentary-informed	80	5.18	0.16	-0.01	0.90
	Commentary-uninformed	76	5.03	0.16	0.01	0.94
	Commentary-misinformed	68	5.31	0.17	0.13	0.28
	Combination-informed	79	5.62	0.16	0.08	0.49
	Combination-uninformed	76	5.00	0.16	0.06	0.61
	Combination-misinformed	61	5.45	0.19	-0.02	0.86

Data from study 1 supported my argument that the impact of different WOM on forecast accuracy depends on the similarity in preferences of source and receiver. Moreover, participants had little insight into the value of WOM for their predictions, and follow-up analyses supported my argument that different forecasting strategies were adopted based on the WOM information available. However, it is conceivable that differences in the products themselves may have confounded the heterogeneity manipulation. Studies 2a and 2b remove any such confounds by examining the moderating influence of source-receiver similarity within the same product experience. In addition, they examine a different category of consumption experience.

2.5 Study 2a

According to my theoretical framework, a critical influence on the value of different WOM for prediction is *source-receiver preference similarity* – i.e., the extent to which reader and reviewer share similar underlying preferences. In particular, being 'matched' with a reviewer whose preferences are similar will dramatically improve the value of ratings, but have little impact on the value of commentary. Therefore, the advantage of commentary for forecasting will diminish (or even disappear) when reader and reviewer have sufficiently similar preferences. Stated formally:

H4: The value of different forms of WOM for forecasting is moderated by source-receiver preference similarity:

a. Commentary information leads to more accurate forecasts than rating information when preference similarity is low, but this advantage diminishes as similarity increases.

b. Combined information leads to more accurate forecasts than rating information when preference similarity is low, but this advantage diminishes as similarity increases.

Study 1 relied on product-level heterogeneity as a proxy for source-receiver similarity, under the assumption that (on average) similarity between randomly paired reviewers and readers will be higher for products with homogenous preferences. In Studies 2a-2b, I examine the preference similarity variable itself. To do so, I employ an experimental design that manipulates preference similarity directly within the same product. Furthermore, in order to examine my theoretical account more closely, I employ textual analysis software to identify specific aspects of commentary that facilitate or inhibit forecasting.

Method

Based on initial pretesting of a variety of music available at Amazon.com, three target music clips were selected to be used as the focal consumption experience. The clips, which were shortened from their original length to 60 seconds, represented three distinct genres (country, jazz, rock); each was unfamiliar to pretest participants, and was neither extremely liked nor disliked. Twenty students at a large university listened to each clip, rated their enjoyment on a 100-pt scale, and wrote a brief commentary. Average enjoyment was as follows: $M_{country}$ = 68.39 (*SD* = 19.13), M_{jazz} = 51.32 (*SD* = 25.32), M_{rock} = 53.33 (*SD* = 25.55). As in study 1, 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*) x 1 (source-receiver similarity) x 3 (music genre: *country* vs. *jazz* vs. *rock*) mixed design. Source-receiver similarity was a continuous measured variable, described below. The WOM type manipulation was identical to that used in study 1, and music genre was treated as a control variable in the analysis.

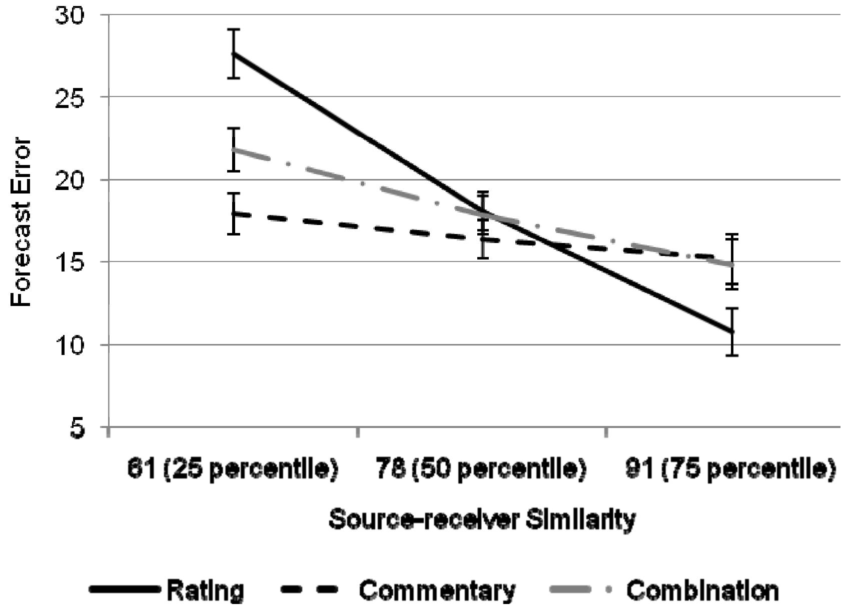
One-hundred sixty-five students participated in the study for course credit. For each of four different music clips (one at a time), participants were asked to read one piece of WOM. Participants then forecasted how much they would enjoy the music on a 100-pt scale. The first clip was a filler not relevant to the study, followed by the three target clips. Presentation of WOM followed the same constrained randomization as study 1, and the same measures were used to assess forecast confidence and perceived reviewer enjoyment.

After making their forecasts, participants were informed that the next task was an unrelated pretest of various pieces of music. All participants 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 liking of various music genres on a 7-pt. scale (-3 = “dislike very much,” 3 = “like very much”); this measure is used in an ancillary analysis below.

Results and Discussion

Actual source-receiver preference similarity was calculated in the same way as study 1. Hypothesis 1 was tested by using a mixed-effect model to predict forecast error as a function of WOM type, source-receiver preference similarity, music genre, and their interactions. Analyses revealed main effects for WOM type ($F(2, 459) = 17.89$, $p < .001$) and source-receiver similarity ($F(1, 459) = 79.54$, $p < .001$). More important, and consistent with predictions, analyses revealed a significant overall interaction between WOM type and source-receiver 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* $F(1, 459) = 16.78$, $p < .001$) (H4a and H4b).

Figure 3 provides a plot of forecast error against an interquartile split of source-receiver preference similarity, and relevant means are presented in Table 5. When similarity was low (below 25th percentile), forecast error in the *commentary* condition ($M = 17.95$) and the *combination* condition ($M = 21.83$) was less than that in the *rating* condition ($M = 27.64$, $F(1, 459) = 23.57$, $p < .001$; $F(1, 459) = 8.88$, $p < .01$). However, this pattern reversed under high source-receiver similarity (above 75th percentile), where forecast error in both *commentary* and *combination* conditions ($M = 15.20$ and 14.86) was greater than that in the *rating* condition ($M = 10.80$, $F(1, 459) = 4.46$, $p = .04$; $F(1, 459) = 3.75$, $p = .05$). These findings support my theoretical framework and demonstrate that source-receiver preference similarity influences the relative value of different WOM for forecasting. Notably, addition of commentary information did not always improve forecasts; when source-receiver similarity was high, *combination* condition forecasters would have been more accurate by simply using the rating provided.



Notes: Forecast error measures the absolute value of the difference between predictions and subsequent evaluations. Means were estimated at the 25th percentile, 50th percentile, and 75th percentile of actual source-receiver preference similarity.

Figure 3: Study 2a: Forecast Error by WOM Type and Source-Receiver Similarity

Table 5: Study 2a: Forecast Error by WOM Type and Actual Source-receiver Similarity / Genre-based Similarity

WOM type	Actual source-receiver similarity: 25th percentile	50th percentile	75th percentile	Genre-based preference similarity: 25th percentile	50th percentile	75th percentile
Rating	27.64	18.09	10.80	22.69	18.70	14.72
	(1.48)	(1.15)	(1.44)	(1.41)	(1.23)	(1.55)
Commentary	17.95	16.39	15.20	17.58	16.89	16.21
	(1.34)	(1.15)	(1.51)	(1.52)	(1.22)	(1.39)
Combination	21.83	17.88	14.86	19.84	19.14	18.44
	(1.27)	(1.18)	(1.52)	(1.41)	(1.24)	(1.51)

Notes: Standard errors are reported in parentheses. Lower forecast error indicates higher accuracy in predicting subsequent evaluations (and thus more useful WOM). Forecast error means were estimated at the 25th percentile, 50th percentile, and 75th percentile of source-receiver similarity and genre-based preference similarity.

A valid concern raised by the use of actual source-receiver similarity is the fact that the receiver's evaluation must be known *a priori*. In the following analysis, I show that my results still hold if the receiver's music genre preferences are used as a proxy for their actual evaluations. I first operationalized genre-based preference similarity for each of the three genres by: 1) transforming the reviewer's evaluation of the clip representing that genre to a 7-pt scale, 2) taking the absolute difference between this value and the participant's 7-pt rating of the genre, and 3) subtracting this difference from 7, so that a higher number indicates greater similarity. Unsurprisingly, a positive correlation was observed between preference similarity based on genre and that based on actual evaluations (r = .44, p <. 001). Consistent with the findings above, results of a mixed-effect analysis using genre-based similarity in place of actual similarity revealed a significant overall interaction of WOM type and preference similarity (F(2, 459) = 5.47, p < .01). In addition, the hypothesized partial interactions were significant: *commentary* vs. *rating* F(1, 459) = 8.25, p < .01; *combination* vs. *rating* F(1, 459) = 8.39, p < .01). Table 3 depicts mean forecast errors, which follow the same pattern obtained in the preceding analysis.

Analysis of the adjustment and confidence measures replicated the findings obtained in study 1. As shown in Table 3, adjustment was lower for the *rating* condition (M = 11.87) than for 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). This finding supports our argument that rating-based forecasts tend to involve only limited adjustment. Furthermore, across all three types of WOM (see Table 4), the correlation between confidence and forecast error was low, indicating little ability to gauge the usefulness of WOM information (see Table 4).

A follow-up analysis was conducted on the textual content of the commentaries provided, in order to identify characteristics of the text that relate to: 1) estimation of the reviewer's evaluation, and 2) inferences of similarity with the reviewer. In the analysis, the 46 commentaries from Study 1 and 60 commentaries from Study 2a were assessed individually with the Linguistic Inquiry and Word Count program (LIWC - Pennebaker, Booth, and Francis 2007). Using a matching algorithm, LIWC searches a target script for words representing 70 linguistic or psychological dimensions, then calculates the percentage of total words in the script that fall into these dimensions. Over numerous investigations, the instrument has been extensively validated and applied to a wide variety of topics and domains, including physical and psychological health, interpersonal relationships, and honesty (Campbell and Pennebaker 2003; Ireland, et al. 2011; Newman, et al. 2003).

My framework suggests that linguistic components of commentary influence the estimation process and the adjustment process separately. In order to investigate the estimation process, I restricted my examination to the *commentary* condition (who did not receive the reviewer's rating directly). The examination took place in three steps. First, *estimation error* was defined as the absolute difference between perceived reviewer enjoyment and the actual enjoyment of the reviewer, 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, correlational analyses were conducted to determine which (if any) linguistic or psychological dimensions of the commentaries predicted their average estimation error. Analyses revealed that, on average, longer reviews did not reduce estimation error (c.f. Mudambi and Schuff 2010) (r = .01, p = .91). However, estimation error was associated with two theoretically relevant LIWC categories. In particular, estimation error was smaller for commentaries that made greater use of *exclusive* words (e.g., 'lack,' 'really,' 'just'; r = -.17, p = .07). This finding is consistent with previous arguments by Pennebaker and King (1999), who suggest that exclusive words help readers to distinguish between possible interpretations of an authors' intended meaning. In addition, 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, a plausible interpretation is that they enable readers to simulate the reviewer's experience more accurately.

In order to investigate the relationship between language use and the adjustment process, I restricted my examination to participants in the *combination* condition. Because these participants received the reviewer's rating, their forecast error is a direct reflection of inaccurate adjustment. The examination proceeded in two steps. First, the average forecast error for each of the 106 commentaries was calculated. Next, in a manner similar to that above, correlation analyses were conducted to determine dimensions of the commentaries that predicted their average forecast error. Analyses revealed that, on average, the overall length of a commentary was not associated with errors in adjustment (r = .06, p = .54). However, adjustment error was associated with two LIWC dimensions that were both theoretically relevant and distinct from those identified above. First, 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). A possible interpretation (corroborated by examination of the reviews) is that the past tense was often used to: 1) describe an experience with the product objectively, which is of no help in gauging similarity, or 2) introspect regarding reasons for their preferences, which can be difficult and misleading (e.g., Wilson and Schooler 1991). In contrast, the future tense was often used to express intentions or provide a context in which they might consume the product again, both of which may be valuable for inferring similarity. Second, adjustment error was larger for commentaries making greater use of function words, including adverbs, pronouns, articles, prepositions, conjunctives, and auxiliary words (r = .19, p = .05). Function words have been described as a '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 attempting to gauge similarity with the reviewer.

In contrast to existing research examining concrete variables such as length, readability, or spelling (Mudambi and Schuff 2010; Ghose and Ipeirotis 2011; Moore 2012), the analyses above focus on the linguistic and psychological dimensions of a commentary. Although the investigation was exploratory, my results suggest that specific, distinct textual properties influence readers' use of the commentary to: 1) estimate the reviewer's evaluation of a product, and 2) adjust away from that evaluation, based on similarity in underlying preferences, 3) make accurate forecasts.

2.6 Study 2b

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

In study 2b, the procedure of study 2a was modified by reversing the order of prediction and evaluation. By this modification, source-receiver similarity could be manipulated *a priori*, and potential dependencies between prediction and evaluation were minimized. Note that 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. To the extent that underlying preferences do not change systematically over the interim, the order of prediction and evaluation is irrelevant. I believe this assumption to be reasonable for music clips, and I 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 expectations and demand effects that were generated by the act of forecasting could not possibly have influenced evaluations. In addition, a time interval of approximately three weeks was introduced between the two stages to ensure that the consumption experience would not influence the later forecasting process.

Because evaluation measures were collected during the first session, actual source-receiver preference similarity could be manipulated directly. Hence, the study incorporated a 3 (WOM type: *rating* vs. *commentary* vs. *combination*) x 2 (preference similarity: *high* vs. *low*) x 3 (music genre: *country* vs. *jazz* vs. *rock*) mixed design. Participants were randomly assigned to one of the six WOM type x similarity conditions, and music genre was a within-subject factor. The WOM type manipulation was identical to that of study 2a. Source-receiver 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 source-receiver 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

A mixed-effect model was used to predict forecast error as a function of WOM type, source-receiver 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 important, and consistent with hypotheses, a significant interaction indicated that the impact of WOM type on forecast error was moderated by similarity ($F(2, 351) = 4.85$, $p < .01$). Both planned partial interaction contrasts were also significant (*commentary* vs. *rating*: $F(1, 351) = 6.20$, $p = .01$; *combination* vs. *rating*: $F(1, 351) = 8.03$, $p < .01$). Results are depicted in Table 2. For participants matched with low-similarity reviewers, forecast error in the *commentary* condition ($M = 22.31$) and *combination* condition ($M = 19.95$) was directionally lower than that in the *rating* condition ($M = 25.63$); 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, the pattern was reversed: 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 ($M = 11.48$, $F(1, 351) = 5.98$, $p = .02$; $F(1, 351) = 3.72$, $p = .06$).

For each condition, Table 3 presents the extent to which participants adjusted their own forecast from their estimate of the reviewer's opinion. As expected, adjustment in the *rating* condition was smaller than adjustment in the *commentary* condition ($M = 13.96$ vs. 19.01, $F(1, 351) = 7.40$, $p < .01$), and marginally smaller than adjustment in the *combination* condition ($M = 16.71$, $F(1, 351) = 2.15$, $p = .14$). As before, the observed correlation between confidence and forecast error (Table 4) did not reflect awareness of the usefulness of different forms of WOM; in fact, the correlation was directionally positive in the *commentary* condition. Moreover, my adjustment measure again suggested that the availability of ratings facilitated anchoring with minimal adjustment, and that this tendency was attenuated when commentary was available.

Taken together, the first three studies provide convergent evidence that neither rating nor commentary has a consistent advantage over the other in aiding prediction. Instead, the relative value of these two types of WOM depends on whether the consumer is paired with a reviewer whose underlying preferences are similar. Because rating alone provides few similarity cues, consumers tend to apply an 'assumed similarity' heuristic which is most effective when source and receiver indeed have similar preferences. On the other hand, consumers given commentary do not have to rely on this heuristic, because preference similarity can be inferred from the commentary itself (albeit imperfectly).

The logic above can be tested by a simple modification to the designs presented thus far: the direct provision of similarity information. For consumers receiving WOM in the form of ratings alone, the addition of similarity information should greatly improve their forecast accuracy. However, the same degree of benefit should not be expected for consumers receiving WOM that contains commentary. Formally:

> **H5:** Explicitly informing consumers about their similarity in preferences to the reviewer will improve forecasts based on ratings alone to a greater extent than forecasts based on commentary or combined information.

2.7 Study 3

The design of study 3 was similar to that of studies 1-2, with two important modifications. First, I included a post-task introspection measure to identify strategies used in processing different types of WOM. Second, participants were provided not only different types of WOM, but also different statements regarding source-receiver preference similarity. Because actual preference similarity was not known in advance, the statements provided randomly were sometimes correct and sometimes incorrect, and this variance provided a straightforward test of my theory. Following the arguments above, forecasts based on a reviewer's rating should benefit strongly from the provision of similarity information, but only when that information is correct. Forecasts based on a reviewer's commentary (with or without a rating) should be relatively unaffected by explicit similarity information, regardless of its accuracy.

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 3 (WOM type: *rating* vs. *commentary* vs. *combination*) x 3 (preference similarity information: *informed* vs. *uninformed* vs. *misinformed*) x 4 (flavor: *root beer* vs. *cinnamon* vs. *pear* vs. *vanilla*) mixed design. Flavor was a within-subject factor, and the WOM type manipulation was the same as that of the prior studies. Preference similarity information represented a within-subject variable, described below.

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 of different flavors (the survey was used as a cover story for the similarity information manipulation). Next, participants were exposed to WOM for each jellybean, one at a time, along with information regarding their preference similarity with the reviewer. Participants in the *informed* and *misinformed* 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, 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 each jellybean, along with their confidence in that forecast, and reported their estimate of the reviewer's enjoyment. Unique to this study was an added manipulation check: participants rated the degree to which they perceived their own preference to be similar to the reviewer's, using a 100-pt scale ("not at all similar," "very similar"). Also added was an introspection measure: at the end of the session, participants were asked to write "a few sentences" describing the process by which they made their forecasts. After a delay of approximately three weeks, participants completed a follow-up session in which they tasted the jellybeans and reported their enjoyment.

Results and Discussion

Examination of the manipulation check revealed that explicit similarity information did influence participants' perceptions of similarity with the reviewer. Compared to the *uninformed* condition, participants told that WOM was provided by a similar peer perceived greater similarity (M = 61.83 vs. 50.44; $F(1, 628) = 26.12$, $p <$ .001); participants told that WOM was provided by a dissimilar peer perceived less similarity (M = 41.98 ($F(1, 628) = 14.37$, $p < .001$).

Among participants who were provided explicit similarity information, the *informed* and *misinformed* conditions were created as follows: For each jellybean, actual source-receiver similarity was calculated in the same manner as studies 1-2. Next, a median split was used to classify participants as either similar or dissimilar to their assigned reviewer, and these classifications were compared to the similarity information provided. The *informed* condition represents participants for whom provided similarity information matched their actual classification (i.e., similar-similar or dissimilar-dissimilar), and the *misinformed* condition represents participants for whom the two did not match.

My framework argues that compared to forecasts based on commentary, those based on ratings will be more heavily influenced by explicit information regarding source-receiver preference similarity. Consistent with this prediction, a mixed-effect analysis revealed that the impact of explicit similarity information on forecast error was moderated by WOM type, as indicated by a significant overall interaction ($F(4, 628)$ = 2.36, $p = .05$). In the *rating* conditions, average forecast error for *uninformed* participants (M = 28.24) was larger than that for *informed* participants (M = 20.68; $F(1, 628) = 4.95$, p = .03), but smaller than that for *misinformed* participants (M = 34.90, $F(1, 628) = 4.09$, p = .04). In the *commentary* and *combination* conditions, however, forecast error was not significantly affected by explicit similarity information (all Fs < 1). These findings support H5 and my argument that participants given ratings alone adjusted their forecasts based on the additional explicit similarity information, while participants given commentary used the content itself to gauge their similarity in preferences with the reviewer.

Additional analyses examined the extent to which adjustment depended on the explicit similarity information provided. For each of the four trials, participants were reclassified by the level of similarity indicated: *similar, dissimilar*, or *control* (no statement given). Next, a mixed-effect model was used to predict adjustment as a function of WOM type, indicated similarity, jellybean flavor, and their interactions. Analysis revealed main effects for WOM type ($F(2, 628) = 11.64$, $p < .001$), indicated similarity ($F(1, 628) = 21.71$, $p < .001$), and flavor ($F(1, 628) = 3.41$, $p = .02$), as well as the expected interaction between WOM type and indicated similarity ($F(3, 628) = 2.38$, $p = .05$). Table 6 summarizes the relevant means. For participants given ratings alone, adjustment was low in both the *control* and *similar* conditions ($M = 12.58$ and $M = 9.43$), and the two were not reliably different ($F(1, 628) < 1$). However, adjustment for both these conditions was lower than adjustment in the *dissimilar* condition ($M = 28.66$, $F(1, 628) = 22.82$, $p < .001$; $F(1, 628) = 30.95$, $p < .001$). These findings support my argument that consumers given ratings alone are especially receptive to explicit similarity information.

In contrast, findings also supported my argument that consumers given commentary will gauge preference similarity based on the content provided. First, a comparison of the three *control* groups revealed adjustment to be significantly higher for the *commentary* and *combination* conditions ($M = 26.82$ and $M = 19.14$) than for the *rating* condition ($M = 12.58$, $F(1, 628) = 18.45$, $p < .001$; $F(1, 628) = 3.93$, $p < .05$), Second, explicit similarity information had limited effect on adjustment when commentary was present. Among the *commentary* groups, adjustment did not reliably differ across *control*, *similar*, and *dissimilar* conditions (M = 26.82, M = 22.25, M = 29.83, NS). Among the *combination* groups, adjustment in the *control* condition (M = 19.14) did not reliably differ from that in the *similar* condition (M = 18.59, NS), but was significantly lower than that in the *dissimilar* condition (M = 29.22, $F(1, 628) = 8.62$, $p < .01$).

Table 6: Study 3: Adjustment by WOM Type and Indicated Similarity

WOM type	Indicated similarity	*n*	*M*	*SD*
Rating	Control	80	12.58	2.31
	High	72	9.43	2.44
	Low	72	28.66	2.45
Commentary	Control	76	26.82	2.37
	High	74	22.25	2.41
	Low	74	29.83	2.41
Combination	Control	76	19.14	2.37
	High	70	18.59	2.47
	Low	70	29.22	2.48

As before, corollary analyses were conducted to examine the correlation between adjustment and actual similarity (Table 3). In the *rating* groups, this correlation was lowest in the *informed* condition and highest in the *misinformed* condition (r = -.30 vs. .39, z = 4.23, p < .001); again suggesting that participants simply took the similarity information at face value. However, in the *commentary* and *combination* groups, the perceived-actual correlation was not reliably different across *informed* and *misinformed* conditions (*commentary*: r = -.28 vs. -.12, NS; *combination*: r = -.47 vs. -.22, NS). The similarity in correlations again suggests that participants formed their own assessments of source-receiver similarity directly from commentary provided.

Finally, participants' verbal reports provided a means of investigating the process by which forecasts were generated. I have argued that the presence of commentary enables consumers to estimate reviewer preferences and make similarity inferences through a process of mental simulation. As a preliminary test of this argument, the content of the verbal 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). Subsequently, an analysis of proportions revealed that reference to simulation was far 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 my framework and identify the use of mental simulation as a factor distinguishing the processing of commentary- and rating-based WOM.

2.8 General Discussion

For the vast majority of consumer decisions, others have already experienced one or more of the options under consideration and shared their own opinions. Growth in e-commerce and communications has enhanced the availability of such word-of-mouth, raising the question of which formats offer the greatest potential for enhancing consumer forecasts. The present research addressed this question by examining two common forms of WOM, numeric ratings and text commentary, and a moderating factor, source-receiver similarity. Consistent with my anchoring-and-adjustment framework, an advantage of commentary over ratings was observed for choices characterized by preference heterogeneity and the inability to match consumers with reviewers having similar preferences. This advantage diminished when preferences were more homogenous, consumers could be matched with similar reviewers, or preference similarity information was provided directly. Together, these findings supplement and integrate prior understanding of word-of-mouth and affective forecasting.

My 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. For example, rating or commentary alone sometimes led to more accurate forecasts than a combination of both (studies 2-3). Moreover, across all studies, I observed low correlations between confidence and forecast accuracy, suggesting a general lack of awareness regarding the value of WOM. Finally, despite the presumed benefits of aggregating reviews via ‘average’ ratings, my study 1 results suggest an important caveat: when opinions of a product vary greatly across consumers, forecasts based on average ratings may underperform those based on a single reviewer (especially if commentary is available).

Theoretical Contributions. Previous work has investigated the influence of various WOM characteristics on persuasion and downstream consequences for sellers (Archak, et al. 2011; Chevalier and Mayzlin 2006); surprisingly, however, the inherent function of WOM as a means of improving consumer decisions has received little attention. I address this void by showing that commonly used forms of WOM may lead to more or less accurate consumer forecasts, depending on the similarity in preferences between source and receiver. An important implication is that despite its utility for vendors, some forms of highly persuasive WOM may generate undesirable outcomes for consumers. Consequently, a number of related questions present themselves: What is the relationship between the persuasiveness of WOM and its objective value as a decision aid? What are the implications of negative versus positive decision outcomes for consumer perceptions of the vendor, reviewer, and WOM more generally? My results suggest that consumers are remarkably unaware of the extent to which to various forms of WOM help or hinder their forecasts, but do consumers learn over time to use WOM information more effectively? Each of these issues merits investigation.

Previous research in affective forecasting (Gilbert, et al. 2009) has demonstrated that the rating of a single peer is often more useful for prediction than descriptive information. I supplement this idea in several ways. First, I compare distinct forms of WOM information from the same source, and propose distinct mechanisms by which they are used for forecasting. Second, I demonstrate that the forecasting advantage of ratings is restricted to cases of high source-receiver preference similarity. Third, I extend these ideas to the concept of preference heterogeneity and find that even when preferences are highly variable, experience-based WOM may nonetheless be valuable, especially when conveyed in the form of commentary.

In exploring the value of ‘average’ ratings for predicting enjoyment, I demonstrated that aggregated WOM information need not outperform that from a single peer. I attributed this result in part to characteristics of consumer decisions that limit the extent to which the “wisdom of crowds” is applicable (Gigone and Hastie 1997). Future work might extend this investigation by focusing on assumptions consumers make regarding their similarity with the ‘average’ reviewer (Cronbach 1955), and how those assumptions are affected by context (e.g., information load, product category) and individual differences (e.g., need-for-uniqueness, thinking style).

Limitations. For the sake of experimental control, my studies incorporated a number of compromises to ecological validity. In particular, all four studies focused on the transmission of WOM from a single source; however, in many typical settings, consumers encounter multiple opinions from multiple sources. Indeed, the acquisition and aggregation of multi-sourced WOM is an important topic unto itself, and although my 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, my research represents only one step towards the development of a more complex, expansive understanding of WOM utilization by consumers.

In all four of my studies, participants were allowed to consider the provided review information 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 my results. A straightforward implication of my anchoring-and-adjustment framework is that load would impede forecasts based on commentary alone to a much 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), I chose to examine product categories that were more hedonic than functional in nature. Compared to hedonic products, functional products tend to evoke less preference heterogeneity; under my framework, this would benefit rating-based forecasts to the extent that the average similarity of reviewers and readers is increased. In addition, functional products tend to be defined by attributes that are tangible and quantifiable; under my framework, this would benefit commentary-based forecasts to the extent that errors of verbalization and simulation are diminished. Hence, I expect that use of functional products would result in generally improved forecasts. More importantly, I believe that the key interaction of similarity and WOM format would operate similarly in a functional setting, but the question remains open.

By design, the present studies provided only sparse objective information about the products involved. Thus, I 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, my studies did not include conditions in which participants received neither ratings nor commentary (such conditions would have provided virtually no basis for prediction). Consequently, I only address the relative performance of ratings and commentary under different levels of source-receiver similarity. Consumers solicit WOM under the assumption that it will aid their decisions; by the use of an appropriate control, it would be interesting to test this assumption directly. Finally, all four studies measured forecasting accuracy based on evaluations; 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 future research using such measures would provide a useful complementary approach.

Implications. The vast majority of web-enabled retailers offer some form of review platform by which consumers may observe the feedback of their peers. Although a broad array of issues must be taken in choosing and implementing such a platform, I suggest that firms carefully consider its 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 on consumer forecasting, and an important underlying consideration is the format of WOM to provide.

From the perspective of my model, the most 'helpful' review is one that transmits an evaluation clearly and provides cues by which readers may accurately infer similarity. Therefore, based on the results of my lexical analyses, consumers might focus on reviews containing qualities associated with better estimation and adjustment (use of affect words, future tense, etc.) More generally, consumers will benefit from knowing whether their preferences are similar to those of the reviewers they encounter. To the extent this information is not provided externally (see below), consumers may seek to obtain it directly. At the individual level, there exist a variety of means by which similarity may be uncovered (e.g., scanning a reviewer's other reviews, his/her profile, or social network account). At the product or category level, consumers may possess lay theories of heterogeneity, based on their own past consumption experiences or knowledge regarding the preferences of their peers. As with the similarity information embedded in a review itself, such tools enable more accurate 'adjustment' from a reviewer's evaluation.

Typically, marketers are aware of the extent to which preferences for a product vary across consumers (distributions of product ratings, prior market research, etc.). My work suggests that the potential advantages of collecting and providing reviewer commentaries will be most pronounced when preferences are known to vary substantially (or when the degree of variance is unknown). In these cases, it is advisable that retailers make salient the availability of commentary information and directly encourage its use. On the other hand, provision of ratings alone may be appropriate for offerings characterized by limited preference heterogeneity. In the latter case, customers would benefit from the presence of cues enabling better inference of source-receiver preference similarity (e.g., a reviewer 'profile'). Because a prospective consumer's evaluation is rarely known *a priori*, various proxies for source-receiver similarity are available: e.g., demographic or psychographic traits, usage characteristics, or evaluations of related products (Naylor, et al. 2011). Given sufficient individual-level data (transaction history, past product reviews, etc.), firms may even be able to approximate the source-receiver similarity of prospective customers with prior reviewers. This process would enable the provision of customized WOM that prioritizes opinions of peers having similar preferences (for example, by arranging reviews in order of 'similarity').

Finally, my work yields insights on a pervasive problem inherent to WOM communication: artificial reviews are common, and a subset of authors with numerous reviews can be overly influential (Kostakos 2009). I suggest that one method of managing the problem relies on the concept of source-receiver preference similarity. Deceitful reviewers are unlikely to be categorized as 'similar' in preferences to any prospective consumer. Therefore, the inclusion of similarity measures in protocols for WOM collection and display provides a potential means of identifying artificial reviews and limiting their undue influence.

CHAPTER 3

WHEN THE CROWD IS DIVIDED: PERCEPTIONS OF WORD-OF-MOUTH DISPERSION

3.1 Introduction

As a direct consequence of advances in information technology, modern consumers increasingly rely upon online word-of-mouth (e.g., product reviews) to guide their purchase decisions. As a result, consumers inevitably encounter a mixture of positive and negative voices for the same products (Berger and Heath 2008; Gershoff, Mukherjee, and Mukhopadhyay 2003). Because uncertainty is typically undesirable in this setting (Mudambi and Schuff 2010), intuition suggests that consumers will tend to favor products with consistent WOM, and existing evidence generally supports this contention (Matz and Wood 2005; Urbany, Dickson, and Wilkie 1989). However, a large proportion of goods and services are characterized by a bimodal distribution of reviewer opinions (Hu, Pavlou, and Zhang 2009), connoting substantial uncertainty and decision risk. The degree to which crowd opinions are divided for a product or service can be described by the dispersion of its WOM distribution. In this research, I examine the influence of WOM dispersion on consumer decisions, and how product and marketing message characteristics affects the process by which dispersion is interpreted.

The topic of mixed opinions has been investigated only sparsely in the extant WOM literature (Cheema and Kaikati 2010; Chevalier and Mayzlin 2006; Irmak, Vallen, and Sen 2010; Naylor, Lamberton, and Norton 2011; Schellekens, Verlegh, and Smidts

2010; Schlosser 2005; Sridhar and Srinivasan 2012) with some notable exceptions. In particular, few scholars have examined the direct impact of different forms of WOM dispersion on consumers' decision making process. Moreover, the few studies that exist have obtained contradictory findings of how greater dispersion might influence purchase decisions—from negative (Moon, Bergey, and Iacobucci 2010; Zhu and Zhang 2010) to positive (Clemons et al. 2006; Martin, Barron, and Norton 2008; Moe and Trusov 2011) or null effects (Chintagunta, Gopinath, and Venkataraman 2010). These conflicting findings suggest potential moderators (Khare et al. 2011; Sun 2011; West and Broniarczyk 1998). In this research, I pursue consolidation by advancing an attribution approach to consumers' interpretation of divided opinions.

Drawing on research in social perception (Boldry and Kashy 1999; Nisbett and Kunda 1985) and attribution (Folkes 1988; Kelley and Michela 1980), I propose that dispersion in WOM for a product will be perceived by consumers as stemming from two general sources: 1) inconsistency in product performance, and 2) idiosyncrasy in preferences across reviewers (preference heterogeneity). I argue that although consumers will frequently attribute WOM dispersion to product-related reasons (#1 above), decision context influences the degree to which dispersion is also attributed to reviewer-related reasons (#2). Moreover, the attribution of WOM dispersion to reviewers rather than the product itself implies that individuals have some control over their product experience, reducing the level of risk implied by mixed opinions. Therefore, I hypothesize that the attribution of WOM dispersion moderates its influence on consumer decisions. In four laboratory studies, I demonstrate the influence of WOM dispersion on purchase intention and provide support for my hypotheses. These findings have important marketing implications for online retailers facing the presentation of mixed product reviews across a variety of categories.

The contributions of this essay are threefold. First, responding to West and Broniarczyk's call for research on the impact of WOM dispersion across different product types (1998), I extend the literature on consumer reactions to this form of information uncertainty. Second, my attribution approach offers a means to consolidate mixed findings in recent WOM literature regarding the influence of WOM variance on consumer behavior. Third, the existing attribution literature offers predictions primarily for settings with high consensus (low dispersion). I supplement this work by providing predictions for settings with low consensus (high dispersion).

3.2 Conceptual Background

Perceptions of Word-of-Mouth Distribution

Psychologists have long been interested in the processes by which individuals perceive social distributions—how different types of attitude or behavior are spread across a particular population (Nisbett and Kunda 1985; Peterson and Beach 1967). In typical examples of this research, participants were asked to estimate the distribution of others' evaluation of movies, food, etc (e.g., How many people will like the movie / be indifferent, not like the movie at all?). These scholars believed that acknowledging the distribution of attitudes and behaviors of others plays an important role in individual decisions, and they found that subject to a few systematic biases, perceived social distributions are often surprisingly accurate.

In contrast, the direct influence of distribution information on decision making has rarely been studied, because the collection and display of information on social distributions was not readily available. However, the flourish of e-commerce and user-generated content has brought social distributions directly into consumers' decision journeys. Technological advancement has allowed individuals to access the opinions of thousands of strangers regarding goods and services in the marketplace: e.g., a consumer can read yelp.com restaurant reviews on their smart phone and decide where to go for dinner, consult imdb.com for movie reviews, and surf cnet.com for electronic reviews. Most contemporary online retail platforms provide entire product rating histories summarized in a bar chart form, making divided opinions easily recognizable (see Figure 4 and 5 for examples). As a result, the distribution of WOM now plays a larger role in the purchase decisions than ever before.

Magic Bullet Express 17-Piece High-Speed Blender Mixing System

by Magic Bullet

Customer Reviews

696 Reviews

5 star:	(231)
4 star:	(139)
3 star:	(66)
2 star:	(82)
1 star:	(178)

Average Customer Review

(696 customer reviews)

Retrieved July 7, 2011, from: <http://www.amazon.com/Magic-Bullet-Express-17-Piece-High-Speed/dp/B000AEZVRS/ref=sr_1_1?ie=UTF8andqid=1290822686andsr=8-1>

Figure 4: Word-of-mouth of Large Dispersion

Retrieved July 7, 2011, from: <http://www.amazon.com/Jurassic-Park-Ultimate-Trilogy-Blu-ray/dp/B0057R5XRG/ref=zg_bs_dvd_8>

Figure 5: Word-of-mouth of Small Dispersion

In consumer research, early WOM literature investigated situations involving consumer consultation with a few friends or family members (Arndt 1967; Brown and Reingen 1987). Only recently have researchers begun to investigate the influence of WOM distributions involving strangers online. Among the fundamental properties of a ratings distribution are its volume, central tendency, and dispersion[2]. Recent empirical studies have examined real-world online product ratings of books, movies, video games, and toiletry products, observing a direct positive effect on sales of volume (Chevalier and Mayzlin 2006; Dellarocas, Zhang, and Awad 2007; Duan, Gu, and Whinston 2008; Li and Hitt 2008; Liu 2006; Moe and Trusov 2011; Sun 2011; Zhu and Zhang 2010) and central tendency (Chevalier and Mayzlin 2006; Chintagunta et al. 2010; Clemons et al. 2006; Dellarocas et al. 2007; Li and Hitt 2008; Moe and Trusov 2011; Moon et al. 2010; Sun 2011; Zhu and Zhang 2010) . Table 7 summarizes relevant findings.

---

[2] I use the statistical definition of dispersion here, which can be operationalized as the variance or second moment of product ratings. This approach follows Clemons et al (2006) but differs from that of others, such as of Godes and Mayzlin (2004). The opposite of dispersion is often known as 'consensus' (Khare, Labrecque, and Asare 2011; West and Broniarczyk 1998).

Table 7: Empirical Literature on WOM Distribution

Article	*Product Category*	*Dependent Variable*	*Characteristic of the Distribution*		
			Volume	*Average*	*Dispersion*
Godes and Mayzlin (2004)	TV shows	TV ratings	No effect		
Chevalier and Mayzlin (2006)	Books	Book sales rank	Positive effect	Positive effect	
Clemons et al. (2006)	Crafted beer	Sales growth rate	No effect	Positive effect	Positive effect
Liu (2006)	Movies	Box office revenue	Positive effect	No effect	
Dellarocas et al. (2007)	Movies	Box office revenue	Positive effect	Positive effect	
Duan et al. (2008)	Movies	Box office revenue	Positive effect	No effect	
Li and Hitt (2008)	Books	Book sales rank	Positive effect	Positive effect	
Chintagunta et al. (2010)	Movies	Box office revenue	No effect	Positive effect	No effect
Moon et al. (2010)	Movies	Box office revenue, satisfaction	Negative effect on satisfaction	Positive effect with ad spending (interaction)	Negative effect on satisfaction
Zhu and Zhang (2010)	Video games	Game sales	Positive effect	Positive effect	Negative effect
Sun (2011)	Books	Book sales rank	Positive effect	Positive effect	Negative effect with high average (interaction)
Moe and Trusov (2011)	Bath, fragrance, and beauty products	Cross-product sales and ratings	Positive effect on sales, negative effect on ratings	Positive effect on sales, negative effect on ratings	Positive effect on sales, negative effect on extreme ratings

In contrast, the influence of dispersion on consumer behavior is less clear, and a small relevant research stream has presented ambiguous conclusions. One the one hand, intuition suggests that consumers may see mixed opinions as a signal of risk—and indeed, greater dispersion of WOM has been shown to lower sales of video games (Zhu and Zhang 2010) and reduce reported satisfaction of movies (Moon et al. 2010). On the other hand, other evidence indicates that mixed opinions are tolerated and even viewed as opportunities (West and Broniarczyk 1998). Clemons, Gao, and Hitt (2006) found that sales of beer brands with mixed WOM grew faster than those with univalent WOM, while Martin, Barron, and Norton (2008) demonstrated that consumers prefer movies with greater WOM dispersion (c.f., Chintagunta et al.(2010), who observed no effects of dispersion on box office revenue). Across a range of hedonic and utilitarian product category.es, Moe and Trusov (2011) observed that higher variance in WOM accompanied sales increases.

Qualifying the Influence of Dispersion

Very few scholars have attempted to consolidate the inconsistency above, and their work has focused on interactions between dispersion and other characteristics of the WOM distribution. Khare et al. (2011) argue that the volume of WOM can enhance existing beliefs: their experimental studies found an effect of WOM dispersion on movie preferences when volume was high (e.g., there are thousands of reviews for a single product), but found no effect when volume was low (e.g., less than a hundred reviews). In contrast, Sun (2011) claims that consumers perceive both product quality and potential undesirable outcomes based on the central tendency and the dispersion of a product's rating distribution. In an examination of consumer ratings and sales data for online book retailers, Sun found that higher ratings variance increased sales only for books with a low average rating is low. A similar pattern would be predicted by West and Broniarczyk's (1998) reference-dependent risk perception theory. Extending concepts from prospect theory (Kahneman and Tversky 1979), West and Broniarczyk argued that consumers' aspiration levels determine their reactions to variance in critic's opinions. In scenarios involving movies and restaurants, participants whose expectations were above the average critic's rating evaluated products more favorably when there was critical disagreement. When participants' expectations were below the average of critics, evaluations were more favorable when there was critical agreement.

Importantly, the investigations above focused almost exclusively on experiential products (books, movies, etc.); little research exists to generalize existing findings to functional products, despite the fact that WOM of functional products can have a distinct influence on consumer search and purchase intention (Huang, Lurie, and Mitra 2009; Senecal and Nantel 2004). Moreover, these approaches cannot fully explain the discrepancies observed in the studies cited above. Given that the existing literature makes no prediction of how consumers perceive WOM dispersion differently across product types and decision domains, the present research aims to fill this gap. This paper employs an attribution-based approach to explore how consumers perceive dispersion. By examining the way variance in reviewer opinions is interpreted, my approach is intended to consolidate the existing mixed findings across different product categories.

Attributions for WOM Dispersion

Attribution theory is predicated on the notion that individuals make spontaneous causal attribution for the events and information that they encounter (Hastie 1984; Kelley 1973; Weiner 1972). Within the marketing literature, consumers' perception of cause-and-effect relationships has been examined extensively in post-purchase contexts (Folkes 1984; Folkes and Kotsos 1986; Hui and Toffoli 2002; McGill 1989; Tsiros, Mittal, and Ross 2004). This stream has shown that attribution of responsibility and stability substantially influences consumer satisfaction judgments, and some scholars have argued that causal inferences about product performance drive most purchase decisions (LeBoeuf and Norton 2012; Weiner 2000). Other work has demonstrated that consumers' causal inferences are not restricted to standalone events such as service encounters, but also the attitudes and behaviors of individuals and groups (Kenworthy and Miller 2002; O'Laughlin and Malle 2002). In keeping with this idea, I adopt an attribution approach to understanding WOM dispersion, suggesting that consumers make different causal inferences for the variance in product ratings under different (and predictable) circumstances.

The attribution process does not happen automatically, and unexpectedness has been identified as one important antecedent to attributional processing (Folkes 1988; Hastie 1984; Kelley and Michela 1980). Scholars have found that people are heavily biased toward perceiving normality in perceiving social distribution shapes; even when the actual distribution lacks this normality (Fried and Holyoak 1984; Nisbett and Kunda 1985). Therefore, highly dispersed WOM distributions, usually characterized by a flat or bimodal form, should instigate unexpectedness and provide a strong motivation to engage in explanatory analysis.

For consumers facing wide dispersion in WOM for a product they are considering, the question becomes: What is the source of dispersion? Prior literature has focused on three potential causal agents: the reviewer, the product, and the situation (Folkes 1988); in a WOM context, the first two of these are most directly applicable. An assumption made by many researchers is that WOM is a proxy for product quality and underlying product characteristics (Khare et al. 2011; Sun 2011; West and Broniarczyk 1998), where product quality is defined as the degree to which a product performs well according to what is advertised. I relax this assumption to suggest that consumers may attribute the variance in product ratings to either the product itself or other factors (e.g., the heterogeneous taste of consumers who provided the WOM, luck, etc.). In fact, negative WOM from a consumer whose opinion diverges from others often leads to the inference that the consumer did not use the product correctly (Laczniak, DeCarlo, and Ramaswami 2001). Moreover, given that reviewers are the contributors of WOM and that characteristics of reviewers often play a role in their interpretation (Forman, Ghose, & Wiesenfeld, 2008; Weiss, Lurie, & MacInnis, 2008), I restrict my focus to causal inferences involving either the product or the reviewers.

My fundamental argument is that the extent to which WOM is attributed to product versus reviewer characteristics will be influenced by consumer expectations regarding how much tastes vary across reviewers. As mentioned in Chapter 2 of the dissertation, consumers hold different beliefs regarding preference heterogeneity—the extent to which preferences for a specific good vary within a population—for different product categories (Gershoff & West, 1998; Price, et al., 1989). When perceived preference heterogeneity is low, a consumer might reasonably expect every reviewer to give the same rating for the same product. In this case, therefore, WOM dispersion should be driven by whether the product consistently delivers what the retailer has promised. For example, faced with electronics that do not receive high evaluations across users, consumers are more likely to assign blame to quality issues than to users who might have operated it incorrectly. In contrast, when perceived preference heterogeneity is high, WOM dispersion can be attributed both to product quality and to idiosyncratic reviewer tastes. For example, variance in ratings for movies, books, and art can be attributed to the product (e.g., the performers, authors, or painter do not entertain as expected.). However, because expectations in these categories are largely subjective (e.g., the same movie trailer may elicit different impressions), product quality will be interpreted differently across consumers. Realizing this, consumers who expect heterogeneity are more likely to attribute WOM dispersion to the reviewers than to the product.

Research shows that products differ by the extent to which consumers share similar preferences, and that people hold beliefs regarding whether the evaluation of a product is similar (Berger and Heath 2007; Gershoff and West 1998; Price, Feick, and Higie 1989). As a general rule, high preference heterogeneity is associated with experiential products such as restaurants and movies (stimuli used by West and Broniarczyk 1998). In contrast, low preference heterogeneity is associated with functional products such as flash drives and desk lamps. (I do not claim that the difference between high and low reviewer preference heterogeneity maps onto the difference between experiential and functional products, which is not the focus of this essay.) As a result, online shoppers are likely to expect reviewer tastes to vary across different product categories. In other words, certain types of products might be more or less conducive to the inference that WOM dispersion is caused by the product versus the reviewer. Next, I explain how these causal attributions interact with context to influence consumer decisions.

Dispersion and Purchase Decision

When WOM dispersion is low, my attribution approach suggests that product type should have little influence purchase decisions. Existing research indicates that individuals tend to expect low dispersion and normality in the distribution of peer opinions (Fried and Holyoak 1984; Nisbett and Kunda 1985). Therefore, WOM with limited variance is hardly surprising, and consumers encountering this WOM have little motivation to make causal inferences (Folkes 1988; Hastie 1984; Kelley and Michela 1980)—in fact, low dispersion indicates invariability in both product- and reviewer-related causes, and may encourage the use of the central tendency as a decision heuristic. In contrast, high dispersion can result from either inconsistent product performance or reviewer preferences, and the salience of one cause or the other will lead to different causal inferences and purchase decisions.

When WOM dispersion is high, the causal inference of that dispersion will affect the decisions made by consumers. I first consider products characterized by low preference heterogeneity (e.g., desk lamps, flash drives) or homogeneous users. Because WOM is often utilized as a proxy for product quality, the dispersion of WOM conveys more information than randomness characterized by lottery tickets or gambles. Thus, this form of dispersion distinguishes itself from forms of 'risk' commonly studied under a prospect theory paradigm (Kahneman and Tversky 1979; West and Broniarczyk 1998). Reduced uncertainty and stability are desirable factors in most consumer decisions (Urbany et al. 1989), and even in experiential domains such as sky diving and mountain climbing, noncontrollable risk is avoided when possible (Celsi, Rose, and Leigh 1993). Therefore, when high WOM dispersion is solely attributed to the product itself, it should have a negative influence on the likelihood of purchase.

However, I argue that the same dispersion is more likely to be tolerated if it belongs to products characterized by high preference heterogeneity (e.g., art, restaurants) or idiosyncratic users. This tolerance has two fundamental reasons: first, internal disagreement is shifted away from the product to external factors like reviewer. Second, attribution of outcomes to a reviewer implies that consumers themselves possess control over the user experience (Lee, Peterson, and Tiedens 2004). Therefore, this type of causal explanation can empower prospective consumers when they read reviews and try to put themselves into the reviewers' shoes (Averill 1973; Hui and Toffoli 2002).

In summary, I hypothesize that although greater WOM dispersion makes consumers less likely to acquire a product, this negative influence of dispersion will be attenuated if the product is characterized by high preference heterogeneity.

H1: The negative influence of high WOM dispersion on consumers' purchase intention will be attenuated if the product is perceived to have high preference heterogeneity.

The attribution approach assumes that consumers make causal attributions during a purchase decision, and that decision contexts affect the direction of these attributions. H1 predicts a pattern in which the context is determined by product type. If the theory is valid, holding product type constant and varying the profile users should lead to similar patterns of results. Specifically, when a product is used by a wide variety of consumers mixed WOM can be expected due to preference heterogeneity, but this is not the case when a product has a narrower user profile. Thus, I predict the following:

H2: The negative influence of high WOM dispersion on consumers' purchase intention will be attenuated if the product is perceived to be used by a diverse group.

3.3 Overview of Studies

My hypotheses were examined in four studies that presented subjects with hypothetical decision scenarios. In all studies, I provide displays of ostensible WOM distributions which included both the overall 'average' rating and the distribution of ratings, using a horizontal bar chart. This format allowed me to vary both the expected outcome of choosing the product (its average rating) and the risk of a obtaining a worse outcome (its dispersion). In Study 1, attribution is manipulated through product preference heterogeneity. Across different levels of average rating, I show that preference heterogeneity moderates the influence of WOM dispersion on purchase intention. Studies 2a and 2b expand my theoretical test by examining choice decisions and incorporating different product types. Study 3 holds product type constant and varies the perceived variability of users, in order to replicate the effect while collecting evidence for the attribution process. Taken together, these studies demonstrate that consumers' reaction to dispersion in product WOM depends critically on the way that dispersion can be attributed. To the extent that preference heterogeneity or user variability is expected, so that dispersion can be attributed to idiosyncratic preferences rather than the product itself, consumers become more tolerant of dispersion.

3.4 Study 1

Study 1 examined the influence of WOM dispersion on purchase intention as a function of product type. In addition, average rating was included in the design, in order to control for the moderating influence of central tendency on dispersion shown by West and Broniarczyk (1998) and Sun (2011). (Note that there is no *a priori* reason to expect that consumer's attribution process will be affected by their aspiration level or the average rating.) Participants expressed their purchase intention under several decision scenarios that varied in WOM. The focal product categories were desk lamps and framed paintings, pretested to represent low and high preference heterogeneity. According to H1, I expected WOM dispersion to have a larger influence on the purchase intention of lamps than paintings.

Method

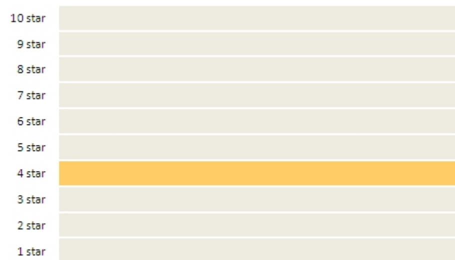
In the main study, a set of sixteen scenarios was presented, one at a time. Each scenario included the same focal product, accompanied by different reviews. The scenarios crossed four levels of average rating with four levels of dispersion. The only between-subject factor, product type, was manipulated by randomly assigning participants to either desk lamps or framed paintings as the target product (see below). Therefore, the main study constituted a 4 (WOM dispersion: *high* vs. *medium* vs. *low* vs. *none*) x 4 (WOM average: *4 star* vs. *5 star* vs. *6 star* vs. *7 star*) x 2 (product type: *desk lamp* vs. *framed painting*) mixed factorial design.

For each of two product types, a core set of eight alternatives was developed and used as stimuli (Figure 6). Two goals were established for this set of alternatives: First, I sought to provide WOM distributions in a format relatively familiar to the participants. Therefore, a star-ratings scale was used to show WOM of product alternatives, with more stars reflecting higher satisfaction; next to each star rating, bars were used to indicate the number of reviewers who have given that rating (see Figure 6). This format is consistent with that of prominent online review sites. Second, in order to control for the potential effect of central tendency on participants' response to dispersion, I sought to vary the average star rating as well as the level of agreement among reviewers. The key independent variable, WOM dispersion, was operationalized as the variance of peer ratings for the target product, and this variable was manipulated (within-subjects) at four levels (high: $var > 8.00$, medium: $var \approx 2.00$, low: $var < 1.00$, and none: $var = 0$). Distributions were created in pairs of high-dispersion and low-dispersion alternatives at four different levels of average rating (4 through 7 out of 10).

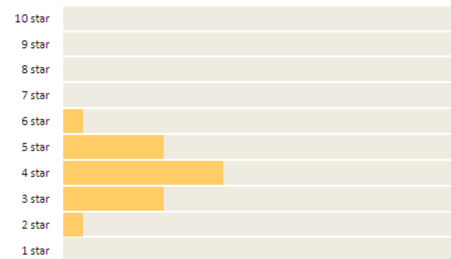
WOM Dispersion (N = 40)

WOM Average	*No dispersion*	*Low dispersion*

4 star

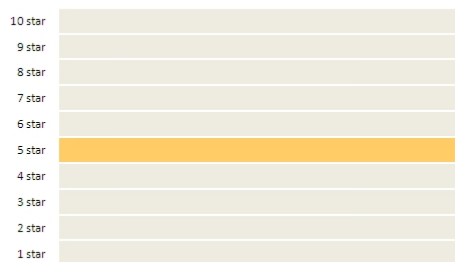


Variance = 0

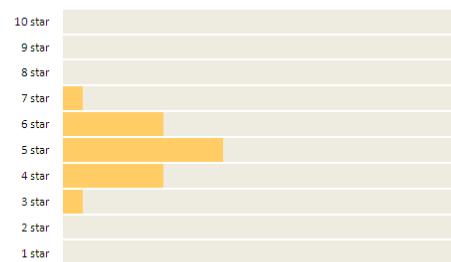


Variance = .9

5 star

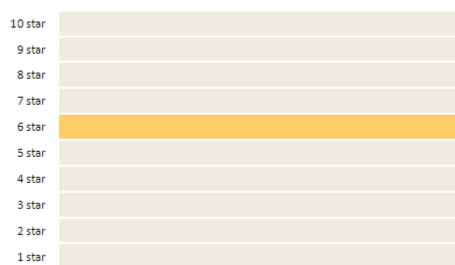


Variance = 0

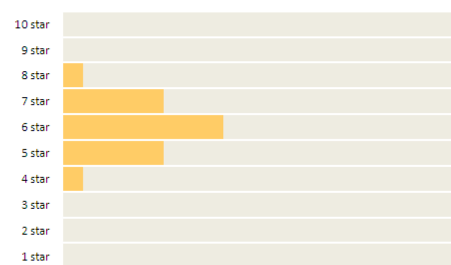


Variance = .9

6 star

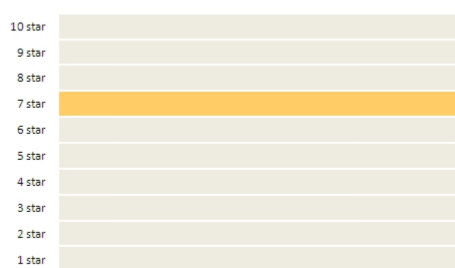


Variance = 0

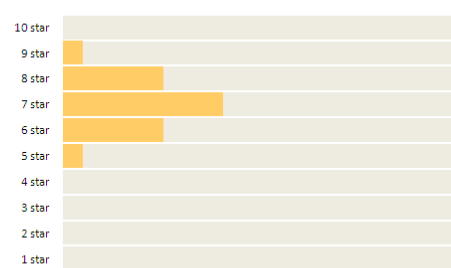


Variance = .9

7 star



Variance = 0



Variance = .9

Notes: Figure 6 continues on the next page.

Figure 6: Study 1: WOM Stimuli

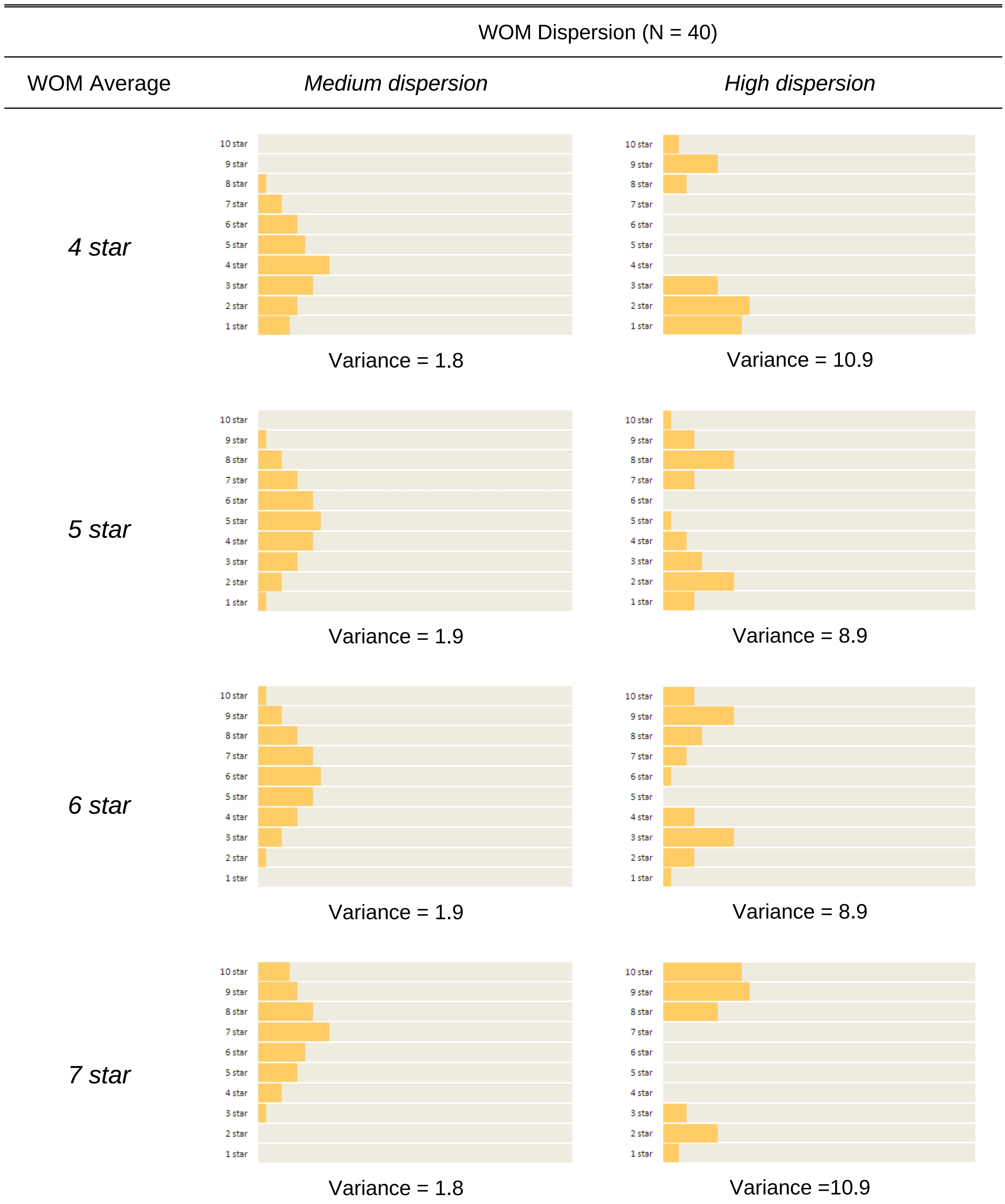


Figure 6 (continued)

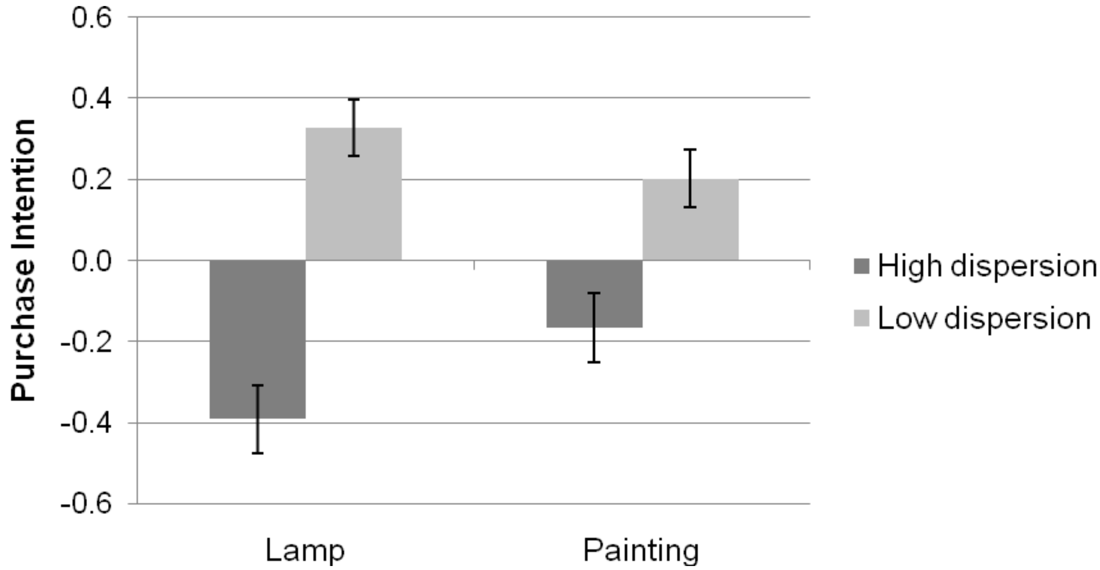
Perceived preference heterogeneity was manipulated by using different types of target product (between-subjects) in the decision scenario. In a pretest, sixty-one subjects were asked to rate several products (e.g., music album, movie, hotel, restaurants, etc.) based on: 1) how similar, and 2) how dissimilar they believe people's evaluations are to one another. Answers to the similarity and dissimilarity questions were combined to form a composite scale of preference heterogeneity beliefs. Based on the results of the pretest, I selected desk lamps for the low-preference-heterogeneity conditions, and framed paintings for the high-preference-heterogeneity conditions (taste similarity for lamps vs. paintings: $M = 69.29$ vs. 34.85 on a 100-pt. scale; $F(1, 60) = 86.60$, $p < .001$).

One-hundred and eighty-nine undergraduate students from the same university participated in the main study in exchange for course credit. Participants were told that they would be evaluating a series of products, based on reviews provided by actual consumers. For each alternative in the set, one at a time, participants saw a screen providing product and WOM information (based on 40 reviewers), then reported their purchase intention on a 9-pt scale (1 = "very unlikely," 9 = "very likely"). The alternatives were shown in a random order.

Results and Discussion

Prior to the analyses, the purchase intention variable was mean-centered to allow for comparisons between the two different product categories. H1 was tested using a mixed–effect model to predict purchase intention as a function of WOM dispersion, WOM average, product type, and all interactions. Analyses revealed main effects of WOM dispersion ($F(3, 1349) = 16.45$, $p < .001$) and WOM average ($F(3, 1352) =$

498.44, $p < .001$). More important, and consistent with H1, analyses revealed a significant overall interaction between dispersion and product type ($F(3, 1349) = 2.55$, $p = .05$), indicating that the influence of dispersion on purchase intention differed by the type of product being considered. A follow-up examination of high and low dispersion conditions revealed that when WOM dispersion was low, participants expressed similar purchase intention for the lamp and painting ($M = .33$ vs. $.20$; $F(1, 744) = 1.52$, $p = .22$), but when dispersion was high, they were less likely to buy the lamp than the painting ($M = -.39$ vs. $-.17$; $F(1, 726) = 3.67$, $p = .06$) (shown in Figure 7). The 3-way interaction was not significant ($F(1, 727) = .58$, $p = .81$), indicating that these results did not depend on the average rating.



Notes: These estimates are mean-adjusted to allow comparison across product types.

Figure 7: Study1: Estimated Means Of Purchase Intention

I also observed an interaction between dispersion and average rating ($F(9, 727) = 7.48$, $p < .001$), indicating that increased dispersion leads to lower purchase intention as average rating increases. This finding replicates prior authors (Sun 2011; West and Broniarczyk 1998) and indicates that dispersion becomes less acceptable once the expected outcome surpasses aspiration levels. In addition, an interaction between average rating and product category ($F(3, 1352) = 5.09$, $p < .01$) was observed, indicating that the average had more influence on intentions for lamps than for paintings. This finding corroborates with the notion that reviews of search goods are perceived as more helpful than reviews of experiential goods (Mudambi and Schuff 2010).

The results of Study 1 supported my argument that the impact of WOM dispersion on purchase intentions depends on preference heterogeneity. I demonstrated that participants preferred low-dispersion options over high-dispersion ones in general, but that compared to the desk lamp decision (low preference heterogeneity), this tendency was attenuated for the framed painting decision (high preference heterogeneity), as participants became more tolerant to high-dispersion options. These findings are consistent with the predictions made by H1, and I show that the moderating effect of preference heterogeneity on the influence of dispersion does not depend on average rating.

Study 1 focused on purchase intention as the dependent measure. The next study examined tradeoffs more directly, using a different dependent measure. The goal was to examine whether my attribution approach can also account for consumer choice.

3.5 Study 2a

Study 2a and 2b expanded the investigation to consumer choices. Because average ratings in online retail settings tend to be positive (Chevalier and Mayzlin 2006; Li and Hitt 2008), and given that H1 was supported across different levels of average rating in Study 1, these studies only included stimuli with moderately positive average ratings. In Study 2a, participants chose between two options that vary by WOM dispersion and average rating. If my theory holds, I should find that participants are more willing to trade high average rating for low dispersion for products of low preference heterogeneity than for products characterized by high preference heterogenity.

Method

Sixty-one paid, nonstudent participants recruited over the internet made a series of choices based on WOM information for nine different product categories. The nine categories included two target categories (desk lamp and framed painting), along with seven fillers. Product preference heterogeneity was manipulated within-subject using the two target products in the decision scenario.

Participants made hypothetical purchase decisions based on online review ratings for products of nine categories. They viewed pairs of alternatives for each of these categories, and were asked to choose between them. Cover stories that specify where these alternatives stand in the market in terms of price and desirability were used to fix participants' aspiration level (West and Broniarczyk 1998) and reduce potential noise in the data. As before, the information included distributions of ratings from ten reviewers for each alternative, represented by a horizontal bar chart. For the two target categories, participants were asked to choose from the following pair of similarly attractive options:

Option A (high-average / high-dispersion): average rating 7/10, *var* = 18.65

Option B (low-average / low-dispersion): average rating 6/10, *var* = .40

The left-right position of alternatives was counterbalanced, and the seven choices were presented in one of two random orderings.

Results and Discussion

Analyses were conducted by comparing the relative choice shares of each option for the two focal categories. This comparison revealed that choices differed reliably across preference heterogeneity conditions ($\chi^2(1) = 3.97$, $p = .05$): specifically, the high-average / high-dispersion Option A was chosen more often for paintings (61%) than for lamps (43%) (see Table 8).

Table 8: Study2a and 2b: Attributes of Options and Choice Share

	Low preference heterogeneity	*High preference heterogeneity*
Study 2a	*Lamp*	*Painting*
Option A (high-average / high-dispersion) *average rating* = 7/10, *var* = 18.65	43%	61%
Option B (low-average / low-dispersion) *average rating* = 6/10, *var* = .40	57%	39%
Study 2b	*Lamp and flash drive*	*Painting and music album*
Option A (high-average / low-dispersion) *average rating* = 6/10, *var* = .40	80%	57%
Option B (low-average / high-dispersion) *average rating* = 5.5/10, *var* = 18.65	20%	43%

This study strengthened the findings of Study 1 by investigating the effects of dispersion and preference heterogeneity in a choice environment, where participants were required to make realistic trade-offs between average ratings and their dispersion. Results of the study supported my argument by showing that more than half of the participants traded lower dispersion for higher average rating for painting (high preference heterogeneity), but less than half did for lamp (low preference heterogeneity).

3.6 Study 2b

My attribution approach suggests that although average ratings may often used as a decision heuristic, WOM dispersion can also influence consumers' decision process. Therefore, it should be possible to find cases where an alternative with a lower WOM average *and* higher WOM dispersion is chosen over an alternative with a higher average *and* lower dispersion. The first alternative represents a high-expected-value, low-risk option, while the second alternative represents a low-expected-value, high-risk option. Expected utility theory suggests that most consumers who are often risk averse would not choose the second alternative. However, my approach argues that for products where high dispersion is attributed to reviewer taste differences rather than product quality, consumers might choose an alternative with lower ratings and higher dispersion.

For the sake of control, previous research examined the influence of variance on choices between two alternatives with same expected values (Meyer 1981; West and Broniarczyk 1998). Much more common in real-world decision making are, choices between alternatives with different expected values. Study 2b examines whether an alternative with lower average rating and higher variance is more likely be chosen in product categories characterized by high preference heterogeneity.

Method

Seventy-nine paid, nonstudent participants recruited over the internet took part in Study 2b. The design and procedure were similar to that of Study 2a, with a few modifications. First, two new target categories were added, based on the pretest described earlier: flash drives (low preference heterogeneity) and downloadable music albums (high preference heterogeneity). In addition, only two filler categories were used, so that participants made a series of choices for six different product categories (four target products and two fillers). Second, the two choices in the target categories were designed so that Option A 'dominated' Option B assuming most people are risk averse (see Figure 8 for an example):

Option A (high-average / low-dispersion): average rating 6/10, *var* = .40

Option B (low-average / high-dispersion): average rating 5.5/10, *var* = 18.65

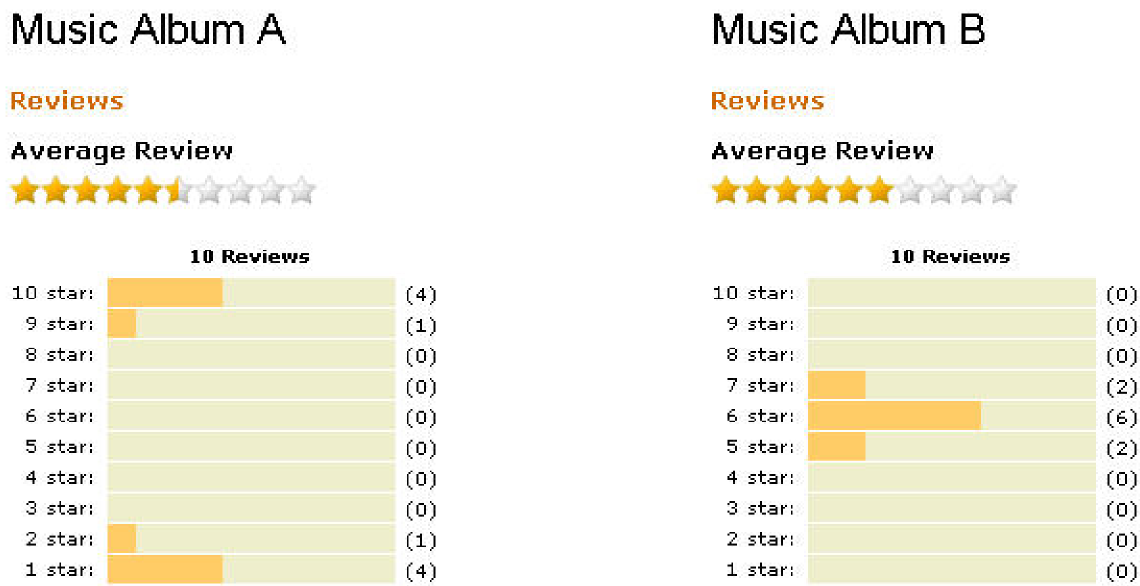


Figure 8: Study 2b: WOM Stimuli Example

Results and Discussion

As before, analyses were conducted by examining the relative choice shares of each option. These analyses again revealed a reliable difference across preference heterogeneity conditions ($\chi^2(1) = 25.51$, $p < .001$). Table 8 summarizes the results. For lamps and flash drives, the 'dominated' option B was chosen only chosen 31 times out of 158 (20%). However, for paintings and music albums, option B was chosen 68 times out of 158 (43%). In other words, option B was chosen by almost half of participants in a product category where preference heterogeneity is expected, despite having both lower expected value and higher variance.

Consistent with my earlier studies, the results of Study 2a demonstrated that the influence of WOM dispersion was affected by product type. In choice scenarios involving products with high preference heterogeneity, even an apparently dominated option became more appealing and was chosen by almost half of the participants. Study 3 extends my investigation by directly testing the role of preference heterogeneity in explaining the effects of dispersion.

3.7 Study 3

The studies conducted thus far tested my attribution approach by manipulating perceived preference heterogeneity through the use of different products. However, these studies are subject to the concern that differences in the products themselves may have been responsible for my results. Therefore, Study 3 was designed to remove potential confounds by holding the product category constant, and instead manipulating the perceived variability of users (described below). According to the logic presented in Hypothesis 2, consumers are more likely to expect a difference in reviewers' evaluations if those reviewers differ by age, gender, profession, etc. As a result, consumers faced with WOM dispersion should be more likely to attribute that dispersion to reviewers when they believe that reviewers come from a diverse group. In addition, Study 3 collected participant attributions directly, in order to examine my underlying process explanation.

Method

One-hundred and thirty-two student participants took part in Study 3 for course credit. The study constituted a 2 (WOM dispersion: *high* vs. *low*) x 2 (user variability: *diverse group* vs. *homogeneous group*) between-subjects factorial design. The focal product was a vacation hotel ('Marison Inn'). Holding the average rating constant (at 6 out of 10 stars), I manipulated WOM dispersion between-subjects through the review summary of forty consumers: high dispersion (*var* = 3.59), low dispersion (*var* = 0.95) (see Figure 9 for examples). User variability was manipulated by telling participants that WOM was posted by one of two groups of reviewers. The *diverse group* conditions stated: "You find some reviews for the hotel on a traveler website populated by an extremely diverse group of customers. The reviewers include business travelers, backpackers, vacationing families, and college students." The *homogenous group* conditions stated: "You find some reviews for the hotel on a traveler website populated by an extremely homogeneous group of customers: college students. The reviewers include college students only."

WOM Dispersion (N = 40)

High dispersion

Variance = 3.6

Low dispersion

Variance = 1.0

Figure 9: Study 3: WOM Stimuli

Participants were asked to imagine that they were traveling to a city across the county and wanted to book a hotel for two nights. A friend informed them of a discount coupon offer in which they could receive \$100 toward their stay for a price of \$50 (i.e., a 50% discount). After seeing reviews about the hotel according to their assigned condition, participants reported their intention to purchase the coupon. (The coupon decision was intended to make participants seriously consider the target option and to reduce potential noise from income differences.) In addition, participants' were asked to provide their expected rating of the hotel, assuming they eventually stayed there. Based on the general principle of self-enhancement biases and the "above-average effect" (Brown 1986; Kruger 1999), consumers who explain WOM dispersion by reviewer differences might expect their own experience to be more satisfactory than the average of reviewers. Finally, direct measures were provided to explore my hypothesized attribution process. Specifically, participants were asked to choose a causal explanation for the observed ratings dispersion, from one of three sources: 'something about the reviewers', 'something about the hotel', or 'other.' (In the analysis, this measure was coded as dichotomous by merging the choice shares of 'reviewer' and 'other.')

Results and Discussion

Analyses were conducted by utilizing ANOVA to predict purchase intention as a function of WOM dispersion, user variability, and their interactions. As before, these analyses revealed a main effect of dispersion ($F(1, 128) = 20.77$, $p < .001$), indicating that more dispersion was associated with lower purchase intention overall. More important, and in line with H2, a dispersion by user variability interaction was obtained

($F(1, 128) = 4.60$, $p < .05$). As shown in Figure 10, the effect of dispersion was smaller when WOM was provided by a diverse group of reviewers than when it was provided by college students only (mean difference = -.61 vs. -1.70). A pairwise comparison showed that the high-variability condition were more likely to buy the hotel coupon than the low-variability condition when exposed to high WOM dispersion ($M = 4.33$ vs. 3.49, $F(1, 128) = 5.37$, $p < .05$). However, This difference disappeared for participants exposed to low WOM dispersion ($M = 4.94$ vs. 5.18, $F(1, 128) <1$, NS).

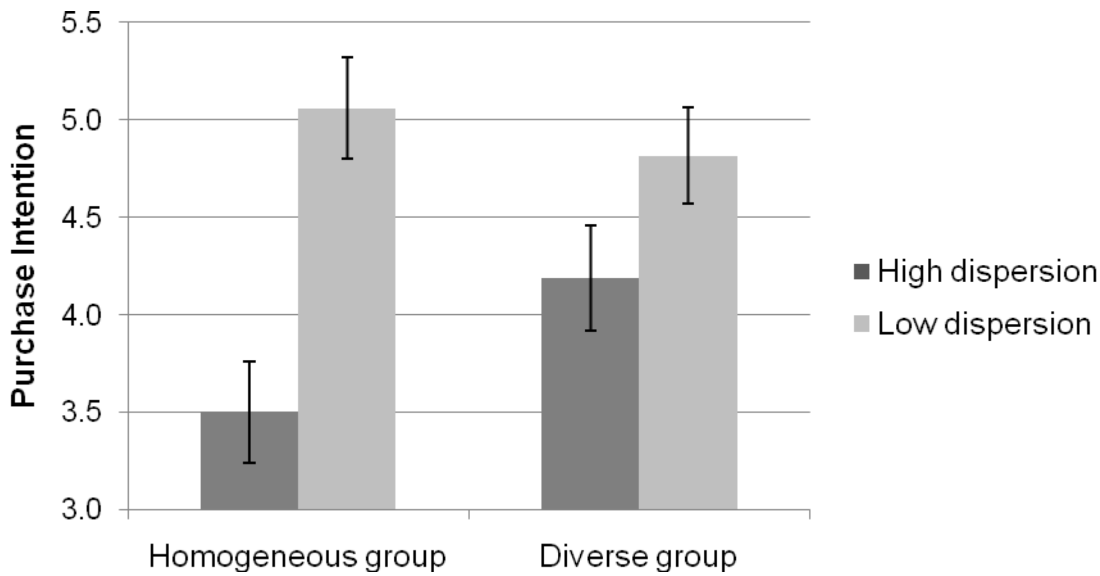


Figure 10: Study3: Estimated Means of Purchase Intention

Participants' own expected rating of the hotel was analyzed in using the same model. Corroborating the findings above, I observed a dispersion by user variability interaction effect ($F(1, 128) = 6.85$, $p = .01$). Again, the high-variability condition rated the hotel coupon higher than the low-variability condition when exposed to high WOM dispersion ($M = 6.80$ vs. 5.52, $F(1, 128) = 17.07$, $p < .001$), and there was no difference for the two conditions exposed to low WOM dispersion ($M = 6.25$ vs. 6.09, $F(1, 128) <1$, NS). As expected, high WOM dispersion led to higher expectations of the hotel experience than low WOM dispersion ($M = 6.80$ vs. 6.25, $F(1, 128) = 3.26$, $p = .07$)—an above-average effect, but only for high-variability conditions. In contrast, high WOM dispersion led to lower-than-average expectations for the low-variability conditions ($M = 5.52$ vs. 6.09, $F(1, 128) = 3.60$, $p = .06$).

Finally, analysis of the direct attribution measure revealed evidence in line with my arguments. When dispersion was high, participants were more likely to attribute it to 'something about the hotel' if they believed that ratings were provided by a homogeneous set of reviewers than by a diverse set of reviewers ($M = 36\%$ vs. 17%; $\chi^2(1) = 3.33$, $p = .07$). In contrast, when dispersion was low, the source of the reviews did not influence attribution ($M = 24\%$ vs. 22%; $\chi^2(1) = .04$, NS).

The results of Study 3 provide evidence for H2 and further support for my attribution approach. When WOM dispersion was high, participants were more willing to acquire a coupon – and expected a better experience – if the group of reviewers was perceived as diverse than if the group of reviewers was perceived as homogeneous. Attribution measures showed that participants' causal explanations were also influenced by the interaction of WOM dispersion and user variability, suggesting that the influence of WOM dispersion on purchase intention depended on the extent to the party to whom it was attributed.

3.8 General Discussion

Modern consumers often consult online WOM before making their purchase decisions. However, when many individuals with different preferences evaluate the same product, the opinions that result are likely to vary. The increasingly common display of WOM as a social distribution not only adds extra information, but also increases risk and uncertainty. Such risk is evident when consumers encounter distributions with high dispersion, which indicates disagreement among reviewers' product usage experiences. Existing research has documented positive as well as negative effects of WOM dispersion on retail performance (e.g., Moe and Trusov 2011; Zhu and Zhang 2010); however, no attempt has been made to consolidate these conflicting results, and we remain largely unaware how conflicting WOM information is processed by consumers psychologically. The present essay is intended to fill these gaps. I argue that perceptions of WOM dispersion are influenced by consumer expectations regarding product preference heterogeneity and user variability. These factors in turn determine how dispersion is attributed and prompt downstream consequences on purchase intention.

Four studies supported my attribution approach to understanding the influence of WOM dispersion. Study 1 provided initial evidence using products of different preference heterogeneity. For products characterized by low preference heterogeneity, options with high WOM dispersion were less attractive to participants. However, for products characterized by high preference heterogeneity, the negative impact with high WOM dispersion was tolerated. Studies 2a and 2b expanded my investigation to choice settings. Consistent with my predictions, options with higher WOM dispersion received greater choice share for products where tastes are expected to vary, and this effect was observed even for a seemingly ‘dominated’ option. Finally, Study 3 held product category constant and manipulated the attribution process through perceived user variability. As expected, the negative influence of increased dispersion was smaller for participants who perceived a diverse user base than for those who perceived a homogeneous user base. Finally, in line with my theoretical arguments, participants in the diverse user conditions were much less likely to attribute WOM dispersion to product-related causes.

Theoretical Contributions. Consumer response to information uncertainty has been investigated extensively, but there are few studies grounded in the emerging WOM context. In particular, West and Broniarczyk (1998) called for research on the investigation of risk embedded in WOM across different product categories. Responding to this call, my essay responds by both demonstrating a moderating effect of product type, and proposing an attribution mechanism to underlie this effect.

Dealing with disagreement in WOM is not a new problem for consumers. Prior research has shown that when WOM received from a limited number of sources lacks consensus, those sources may be treated differently based on their credibility or diagnosticity. For example, WOM from a similar peer—or a consistent critic—may be weighted more heavily (West and Broniarczyk 1998). However, the selective strategy described above is simply not feasible in most contemporary online environments, where

WOM exists in vast amount and comes from unknown others. Therefore, consumers must resort to other information processing strategies. Here, I suggest that expectancies regarding preference heterogeneity (and resulting attributions) play an important role in those strategies.

Previous work in WOM has found seemingly contradictory associations between review variance and sales. The attribution approach proposed in this essay offers a way to consolidate these mixed findings. For example, a positive influence of WOM dispersion has been found for toiletry products with a variety of fragrances (Moe and Trusov 2011) and for craft beers (Clemons et al. 2006); both of these appear to represent product categories characterized by high preference heterogeneity. In contrast, a negative effect or null effects have been observed for movies (Chintagunta et al. 2010) and video games (Zhu and Zhang 2010). In these categories, it is likely that reviewers are perceived as relatively homogenous in tastes, so that their variance of opinion creates uncertainty as to whether the product can deliver what is promised.

Within consumer research, the relevant attribution literature has often focused on causal inferences regarding single pieces of discrepant WOM (Folkes 1988; Laczniak, DeCarlo, and Ramaswami 2001). In contrast, I explore the attributions made by consumers for multiple, discrepant WOM sources and demonstrate the influence of those attributions on purchase decisions. Therefore, while extant research has explored settings marked by high consensus (low WOM dispersion), I supplement this work by examining settings where no consensus exists (high WOM dispersion).

Managerial Implications. The findings of this essay have important implications for marketers, especially those facing divided consumer opinions of their offerings.

Intuition suggests that the risk conveyed by polarized WOM tends to drive away prospective customers. Using my attribution approach, I point out that this may not be true. In particular, the negative influence of WOM dispersion varies by product category. Therefore, for marketers of products where tastes are expected to vary or those with a very diverse user base, it may be reasonable to worry less about divided opinions.

This essay also yields insights on targeting and branding. Practitioners are often advised to concentrate their efforts on specific subsets of consumers for whom the value proposition is strongest. However, in settings characterized by pervasive WOM, my research suggests that fine-grained targeting may be problematic, especially when WOM dispersion is high. A clear implication of Study 3 is that if a product is positioned narrowly, the presence of mixed opinions can lead to inferences of inconsistent quality and lower intentions to purchase. However, if a product is positioned to appeal more broadly, variance in product ratings will tend to be attributed to user idiosyncrasies and therefore tolerated. Above the actual targeting decision, various communication tactics can be used to signal a wide or narrow user base, having similar effects (e.g., the use of testimonials from a wide assortment of consumers to signal diversity). More generally, by considering the influence of their own actions on the attribution process, marketers can practically influence the choices of consumers facing divided WOM.

REFERENCES

Adaval, Rashmi and Robert S. Wyer Jr (1998), "The Role of Narratives in Consumer Information Processing," *Journal of Consumer Psychology*, 7 (3), 207.

Ansari, Asim and Carl F. Mela (2003), "E-Customization," *Journal of Marketing Research*, 40 (2), 131-45.

Archak, Nikolay, Anindya Ghose, and Panagiotis G. Ipeirotis (2011), "Deriving the Pricing Power of Product Features by Mining Consumer Reviews," *Management Science*, forthcoming.

Arndt, Johan (1967), "Role of Product-Related Conversations in the Diffusion of a New Product," *Journal of Marketing Research*, 4 (3), 291-95.

Averill, James R. (1973), "Personal Control over Aversive Stimuli and Its Relationship to Stress," *Psychological Bulletin*, 80 (4), 286-303.

Berger, Jonah and Chip Heath (2007), "Where Consumers Diverge from Others: Identity Signaling and Product Domains," *Journal of Consumer Research*, 34 (2), 121-34.

Berger, Jonah and Chip Heath (2008), "Who Drives Divergence? Identity Signaling, Outgroup Dissimilarity, and the Abandonment of Cultural Tastes," *Journal of Personality and Social Psychology*, 95 (3), 593-607.

Berlo, David K. (1960), *The Process of Communication*, San Francisco: Holt, Rinehart, Winston.

Billeter, Darron, Ajay Kalra, and George Loewenstein (2011), "Underpredicting Learning after Initial Experience with a Product," *Journal of Consumer Research*, 37 (5), 723-36.

Boldry, Jennifer G. and Deborah A. Kashy (1999), "Intergroup Perception in Naturally Occurring Groups of Differential Status: A Social Relations Perspective," *Journal of Personality and Social Psychology*, 77 (6), 1200-12.

Borgida, Eugene and Richard. E. Nisbett (1977), "The Differential Impact of Abstract vs. Concrete Information on Decisions," *Journal of Applied Social Psychology*, 7 (3), 258-71.

Brown, Jacqueline Johnson and Peter H. Reingen (1987), "Social Ties and Word-of-Mouth Referral Behavior," *Journal of Consumer Research*, 14 (3), 350-62.

Brown, Jonathon D. (1986), "Evaluations of Self and Others: Self-Enhancement Biases in Social Judgments," *Social Cognition*, 4 (4), 353-76.

Celsi, Richard L., Randall L. Rose, and Thomas W. Leigh (1993), "An Exploration of High-Risk Leisure Consumption through Skydiving," *Journal of Consumer Research*, 20 (1), 1-23.

Cheema, Amar and Andrew M. Kaikati (2010), "The Effect of Need for Uniqueness on Word of Mouth," *Journal of Marketing Research*, 47 (3), 553-63.

Chevalier, Judith A. and Dina Mayzlin (2006), "The Effect of Word of Mouth on Sales: Online Book Reviews," *Journal of Marketing Research*, 43 (3), 345-54.

Chintagunta, Pradeep K., Shyam Gopinath, and Sriram Venkataraman (2010), "The Effects of Online User Reviews on Movie Box Office Performance: Accounting for Sequential Rollout and Aggregation across Local Markets," *Marketing Science*, 29 (5), 944-57.

Clemons, Eric K., Guodong Gordon Gao, and Lorin M. Hitt (2006), "When Online Reviews Meet Hyperdifferentiation: A Study of the Craft Beer Industry," *Journal of Management Information Systems*, 23 (2), 149-71.

Cronbach, Lee (1955), "Processes Affecting Scores on "Understanding of Others" and "Assumed Similarity."," *Psychological Bulletin*, 52 (3), 177-93.

Davis, Harry L., Stephen J. Hoch, and E. K. Easton Ragsdale (1986), "An Anchoring and Adjustment Model of Spousal Predictions," *Journal of Consumer Research*, 13 (1), 25-37.

Dellarocas, Chrysanthos, Xiaoquan Zhang, and Neveen F. Awad (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, Peter R. (1982), "The Impact of Enriching Case and Statistical Information on Consumer Judgments," *Journal of Consumer Research*, 8 (4), 398-406.

Duan, Wenjing, Bin Gu, and Andrew B. Whinston (2008), "The Dynamics of Online Word-of-Mouth and Product Sales—an Empirical Investigation of the Movie Industry," *Journal of Retailing*, 84 (2), 233-42.

Dunning, D., Griffin, D. W., Milojkovic, J. D., and Ross L. (1990), "The Overconfidence Effect in Social Prediction," *Journal of Personality and Social Psychology*, 58(4), 568-581.

eMarketer. (2011), "Negative Buzz Gains Traction Among Web Users," Retrieved September 28, 2011, from: http://www.emarketer.com/Article.aspx?R=1008614

Folkes, Valerie S. (1984), "Consumer Reactions to Product Failure: An Attributional Approach," *Journal of Consumer Research*, 10 (4), 398-409.

Folkes, Valerie S. (1988), "Recent Attribution Research in Consumer Behavior: A Review and New Directions," *Journal of Consumer Research*, 14 (4), 548-65.

Folkes, Valerie S. and Barbara Kotsos (1986), "Buyers' and Sellers' Explanations for Product Failure: Who Done It?," *Journal of Marketing*, 50 (2), 74-80.

Forman, Chris, Anindya Ghose, and Batia Wiesenfeld (2008), "Examining the Relationship Between Reviews and Sales: The Role of Reviewer Identity Disclosure in Electronic Markets," *Information Systems Research*, 19 (3), 291-313.

Freedman, L. (2008), *Merchant and Customer Perspectives on Customer Reviews and User-generated Content*.

Fried, Lisbeth S. and Keith J. Holyoak (1984), "Induction of Category Distributions: A Framework for Classification Learning," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10 (2), 234-57.

Gershoff, Andrew D., Susan M. Broniarczyk, and Patricia M. West (2001), "Recommendation or Evaluation? Task Sensitivity in Information Source Selection," *Journal of Consumer Research*, 28 (3), 418-38.

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

Gershoff, Andrew D., Ashesh Mukherjee, and Anirban Mukhopadhyay (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, Andrew D. and Patricia M. West (1998), "Using a Community of Knowledge to Build Intelligent Agents," *Marketing Letters*, 9 (1), 79-91.

Gigone, Daniel and Reid Hastie (1997), "Proper Analysis of the Accuracy of Group Judgments," *Psychological Bulletin*, 121 (1), 149-67.

Gilbert, Daniel T., Matthew A. Killingsworth, Rebecca N. Eyre, and Timothy D. Wilson (2009), "The Surprising Power of Neighborly Advice," *Science*, 323 (5921), 1617-19.

Gilbert, Daniel T. and Timothy D. Wilson (2007), "Prospection: Experiencing the Future," *Science*, 317 (5843), 1351-54.

Godes, David and Dina Mayzlin (2004), "Using Online Conversations to Study Word-of-Mouth Communication," *Marketing Science*, 23 (4), 545-60.

Godes, David and Dina Mayzlin (2009), "Firm-Created Word-of-Mouth Communication: Evidence from a Field Test," *Marketing Science*, 28 (4), 721-39.

Godes, David and José Silva (2009), "The Dynamics of Online Opinion," Working Paper.

Hastie, Reid (1984), "Causes and Effects of Causal Attribution," *Journal of Personality and Social Psychology*, 46 (1), 44-56.

Hoch, Stephen J. (1988), "Who Do We Know: Predicting the Interests and Opinions of the American Consumer," *Journal of Consumer Research*, 15 (3), 315-24.

Hu, Nan, Paul A. Pavlou, and Jie Zhang (2009), "Overcoming the J-shaped Distribution of Product Reviews," *Communications of the ACM*, 52 (10), 144-47.

Huang, Peng, Nicholas H. Lurie, and Sabyasachi Mitra (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.

Hui, Michael K. and Roy Toffoli (2002), "Perceived Control and Consumer Attribution for the Service Encounter," *Journal of Applied Social Psychology*, 32 (9), 1825-44.

Irmak, Caglar, Beth Vallen, and Sankar Sen (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), 443-55.

Kahneman, Daniel and Jackie Snell (1992), "Predicting a Changing Taste: Do People Know What They Will Like?" *Journal of Behavioral Decision Making*, 5 (3), 187-200.

Kahneman, Daniel and Amos Tversky (1979), "Prospect Theory: An Analysis of Decision under Risk," *Econometrica*, 47 (2), 263-91.

Kahneman, Daniel and Amos Tversky (1982), "The Simulation Heuristic," in *Judgment under Uncertainty: Heuristics and Biases*, ed. Daniel Kahneman, Paul Slovic and Amos Tversky, Cambridge: Cambridge University Press.

Kelley, Harold H. (1973), "The Processes of Causal Attribution," *American Psychologist*, 28 (2), 107-28.

Kelley, Harold H. and John L. Michela (1980), "Attribution Theory and Research," *Annual Review of Psychology*, 31 (1), 457-501.

Kenworthy, Jared B. and Norman Miller (2002), "Attributional Biases About the Origins of Attitudes: Externality, Emotionality and Rationality," *Journal of Personality and Social Psychology*, 82 (5), 693-707.

Khare, Adwait, Lauren I. Labrecque, and Anthony K. Asare (2011), "The Assimilative and Contrastive Effects of Word-of-Mouth Volume: An Experimental Examination of Online Consumer Ratings," *Journal of Retailing*, 87 (1), 111-26.

Kostakos, V. (2009), "Is the Crowd's Wisdom Biased? A Quantitative Analysis of Three Online Communities," Vol. 4: IEEE, 251-55.

Kruger, Justin (1999), "Lake Wobegon Be Gone! The "Below-Average Effect" and the Egocentric Nature of Comparative Ability Judgments," *Journal of Personality and Social Psychology*, 77 (2), 221-32.

Laczniak, Russell N., Thomas E. DeCarlo, and Sridhar N. Ramaswami (2001), "Consumers' Responses to Negative Word-of-Mouth Communication: An Attribution Theory Perspective," *Journal of Consumer Psychology*, 11 (1), 57-73.

Larrick, Richard P. and Jack B. Soll (2006), "Intuitions About Combining Opinions: Misappreciation of the Averaging Principle," *Management Science*, 52 (1), 111-27.

LeBoeuf, Robyn A. and Michael I. Norton (2012), "Consequence-Cause Matching: Looking to the Consequences of Events to Infer Their Causes," *Journal of Consumer Research*, 39 (1), 128-41.

Lee, Fiona, Christopher Peterson, and Larissa Z. Tiedens (2004), "Mea Culpa: Predicting Stock Prices from Organizational Attributions," *Personality and Social Psychology Bulletin*, 30 (12), 1636-49.

Li, Xinxin and Lorin M. Hitt (2008), "Self-Selection and Information Role of Online Product Reviews," *Information Systems Research*, 19 (4), 456-74.

Lichtenstein, Sarah, Paul Slovic, Baruch Fischhoff, Mark Layman, and Barbara Combs (1978), "Judged Frequency of Lethal Events," *Journal of Experimental Psychology: Human Learning and Memory*, 4 (6), 551-78.

Liu, Yong (2006), "Word of Mouth for Movies: Its Dynamics and Impact on Box Office Revenue," *Journal of Marketing*, 70 (3), 74-89.

Loewenstein, George and Daniel Adler (1995), "A Bias in the Prediction of Tastes," *Economic Journal*, 105 (431), 929-37.

Loewenstein, George and David Schkade (1999), "Wouldn't It Be Nice? Predicting Future Feelings," in *Well-being: The foundations of hedonic psychology*, ed. Daniel Kahneman, Ed Diener and Norbert Schwarz, New York: Russell Sage Foundation, 85.

Martin, Jolie M., Gregory M. Barron, and Michael I. Norton (2008), "Response to Variance in the Opinions of Others: Preferable in Positive Domains, Aversive in Negative Domains," *Society for Consumer Psychology Conference Proceedings*, New Orleans, LA.

Matz, David C. and Wendy Wood (2005), "Cognitive Dissonance in Groups: The Consequences of Disagreement," *Journal of Personality and Social Psychology*, 88 (1), 22-37.

McGill, Ann L. (1989), "Context Effects in Judgments of Causation," *Journal of Personality and Social Psychology*, 57 (2), 189-200.

Mellers, Barbara A., Alan Schwartz, Katty Ho, and Ilana Ritov (1997), "Decision Affect Theory: Emotional Reactions to the Outcomes of Risky Options," *Psychological Science*, 8 (6), 423-29.

Meyer, Robert J. (1981), "A Model of Multiattribute Judgments under Attribute Uncertainty and Informational Constraint," *Journal of Marketing Research*, 18 (4), 428-41.

Moe, Wendy W. and Michael Trusov (2011), "The Value of Social Dynamics in Online Product Ratings Forums," *Journal of Marketing Research*, 48 (3), 444-56.

Moon, Sangkil, Paul K. Bergey, and Dawn Iacobucci (2010), "Dynamic Effects among Movie Ratings, Movie Revenues, and Viewer Satisfaction," *Journal of Marketing*, 74 (1), 108-21.

Moore, Sarah G. (2012), "Some Things Are Better Left Unsaid: How Word of Mouth Influences the Storyteller," *Journal of Consumer Research*, 38 (6), 1140-54.

Mudambi, Susan M. and David Schuff (2010), "What Makes a Helpful Online Review? A Study of Customer Reviews on Amazon.Com," *MIS Quarterly*, 34 (1), 185-200.

Myers, David G. (1998), *Psychology*, 5th ed., New York: Worth.

Naylor, Rebecca Walker, Cait Poynor Lamberton, and David A. Norton (2011), "Seeing Ourselves in Others: Reviewer Ambiguity, Egocentric Anchoring, and Persuasion," *Journal of Marketing Research*, 48 (3), 617-31.

Nisbett, Richard E. and Ziva Kunda (1985), "Perception of Social Distributions," *Journal of Personality and Social Psychology*, 48 (2), 297-311.

Nelson, Phillip (1970), "Information and Consumer Behavior," *The Journal of Political Economy*, 78 (2), 311-29.

O'Laughlin, Matthew J. and Bertram F. Malle (2002), "How People Explain Actions Performed by Groups and Individuals," *Journal of Personality and Social Psychology*, 82 (1), 33-48.

Olshavsky, Richard W. and John A. Miller (1972), "Consumer Expectations, Product Performance, and Perceived Product Quality," *Journal of Marketing Research*, 9 (1), 19-21.

Park, Do-Hyung, Jumin Lee, and Ingoo Han (2007), "The Effect of On-Line Consumer Reviews on Consumer Purchasing Intention: The Moderating Role of Involvement," *International Journal of Electronic Commerce*, 11 (4), 125-48.

Patrick, Vanessa M., Deborah J. MacInnis, and C. Whan Park (2007), "Not as Happy as I Thought I'd Be? Affective Misforecasting and Product Evaluations," *Journal of Consumer Research*, 33 (4), 479-89.

Peterson, Cameron R. and Lee R. Beach (1967), "Man as an Intuitive Statistician," *Psychological Bulletin*, 68 (1), 29-46.

Price, Linda L., Lawrence F. Feick, and Robin A. Higie (1989), "Preference Heterogeneity and Coorientation as Determinants of Perceived Informational Influence," *Journal of Business Research*, 19 (3), 227-42.

Read, Daniel and George Loewenstein (1995), "Diversification Bias: Explaining the Discrepancy in Variety Seeking Between Combined and Separated Choices," *Journal of Experimental Psychology: Applied*, 1 (1), 34-49.

Rothwell, D. J. (2010), *In the Company of Others: An Introduction to Communication*, (3rd ed.). New York: Oxford.

Schellekens, Gaby A. C., Peeter W. J. Verlegh, and Ale Smidts (2010), "Language Abstraction in Word of Mouth," *Journal of Consumer Research*, 37 (2), 207-23.

Schlosser, Ann E. (2005), "Posting Versus Lurking: Communicating in a Multiple Audience Context," *Journal of Consumer Research*, 32 (2), 260-65.

Schlosser, Ann 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, Jonathan W. and Tonya Y. Engstler-Schooler (1990), "Verbal Overshadowing of Visual Memories: Some Things Are Better Left Unsaid," *Cognitive Psychology*, 22 (1), 36-71.

Senecal, Sylvain and Jacques Nantel (2004), "The Influence of Online Product Recommendations on Consumers' Online Choices," *Journal of Retailing*, 80 (2), 159-69.

Sengupta, Jaideep and Gavan J. Fitzsimons (2000), "The Effects of Analyzing Reasons for Brand Preferences: Disruption or Reinforcement?" *Journal of Marketing Research*, 37(3), 318-30.

Shafir, Eldar, Itamar Simonson, and Amos Tversky (1993), "Reason-Based Choice," *Cognition*, 49(1), 11-36.

Shiv, Baba and Joel Huber (2000), "The Impact of Anticipating Satisfaction on Consumer Choice," *Journal of Consumer Research*, 27 (2), 202-16.

Simonson, Itamar (1990), "The Effect of Purchase Quantity and Timing on Variety-Seeking Behavior," *Journal of Marketing Research*, 27 (2), 150-62.

Sridhar, Shrihari and Raji Srinivasan (2012), "Social Influence Effects in Online Product Ratings," *Journal of Marketing*, 1-49.

Sun, Monic (2011), "How Does Variance of Product Ratings Matter?," *Management Science*.

Tsiros, Michael, Vikas Mittal, and William T Ross, Jr. (2004), "The Role of Attributions in Customer Satisfaction: A Reexamination," *Journal of Consumer Research*, 31 (2), 476-83.

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

Urbany, Joel E., Peter R. Dickson, and William L. Wilkie (1989), "Buyer Uncertainty and Information Search," *Journal of Consumer Research*, 16 (2), 208-15.

Wang, Jing, Nathan Novemsky, and Ravi Dhar (2009), "Anticipating Adaptation to Products," *Journal of Consumer Research*, 36 (2), 149-59.

Watts, Duncan J and Peter Sheridan Dodds (2007), "Influentials, Networks, and Public Opinion Formation," *Journal of Consumer Research*, 34(4), 441-458.

Weiner, Bernard (1972), "Attribution Theory, Achievement Motivation, and the Educational Process," *Review of Educational Research*, 42 (2), 203-15.

Weiner, Bernard (2000), "Attributional Thoughts About Consumer Behavior," *Journal of Consumer Research*, 27 (3), 382-87.

Weiss, Allen M., Nicholas H. Lurie, and Deborah J. MacInnis (2008), "Listening to Strangers: Whose Responses Are Valuable, How Valuable Are They, and Why?," *Journal of Marketing Research*, 45 (4), 425-36.

West, Patricia M. and Susan M. Broniarczyk (1998), "Integrating Multiple Opinions: The Role of Aspiration Level on Consumer Response to Critic Consensus," *Journal of Consumer Research*, 25 (1), 38-51.

Wilson, Timothy D. and Jonathan W. Schooler (1991), "Thinking Too Much: Introspection Can Reduce the Quality of Preferences and Decisions," *Journal of Personality and Social Psychology*, 60 (2), 181-92.

Wilson, Timothy D., Thalia Wheatley, Jonathan M. Meyers, Daniel T. Gilbert, and Danny Axsom (2000), "Focalism: A Source of Durability Bias in Affective Forecasting," *Journal of Personality and Social Psychology*, 78 (5), 821-36.

Wood, Stacy L. and James R. Bettman (2007), "Predicting Happiness: How Normative Feeling Rules Influence (And Even Reverse) Durability Bias," *Journal of Consumer Psychology*, 17 (3), 188-201.

Zhao, Min, Steve Hoeffler, and Gal Zauberman (2007), "Mental Simulation and Preference Consistency over Time: The Role of Process- Versus Outcome-Focused Thoughts," *Journal of Marketing Research*, 44 (3), 379-88.

Zhu, Feng and Xiaoquan Zhang (2010), "Impact of Online Consumer Reviews on Sales: The Moderating Role of Product and Consumer Characteristics," *Journal of Marketing*, 74 (2), 133-48.