Seeing the Forest for the Trees: Information Aggregation in Online Decision-Making

A dissertation presented

by

Jolie Mae Martin

to

The Committee for the PhD in Science, Technology, and Management

in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy
in the subject of
Science, Technology, and Management

Harvard University
Cambridge, Massachusetts

May, 2008

Professor Michael I. Norton Jolie Mae Martin

Seeing the Forest for the Trees: Information Aggregation in Online Decision-Making

Abstract

With growing reliance on the Internet as a primary source of input into nearly every type of decision, individuals must find ways to cope with the overwhelming quantity and variety of information accessible to them. Unfortunately, while technology increases the availability of information by facilitating both sharing and retrieval, it does not yet offer clear assistance for integrating this information into coherent preferences. Furthermore, the weighting of disparate information content is subject to the inherent decision-making biases that people have been shown to exhibit in many other contexts, and reliance on simplifying heuristics may even be exacerbated in online environments where distillation of meaning from abundant or conflicting information is especially difficult.

The first paper looks at the effect of interface design on decisions, whereby individuals focus their attention according to the organization of information. Simply manipulating the way that option attributes are partitioned into categories, we induce decision-makers to ascribe different relative importance to them. The second paper

examines the interpretation of opinion information, and an observed asymmetric preference for high variance experiences in positive domains and low variance experiences in negative domains. We argue that salient memories of prior experiences set reference points at extremes rather than “null” outcomes, and in turn, decision-makers perceive disproportionate likelihood and impact of realizing highly favorable outcomes in positive domains and highly unfavorable outcomes in negative domains. Finally, the third paper explores unrepresentativeness of opinion information that is made public, as a result of sample bias in the choice to provide ratings and subsequent response bias in the ratings provided. Specifically, we demonstrate that individuals are more inclined to share extreme opinions, and the opinions they do share are influenced by exposure to the opinions of others.

In each of the experimental studies described, we highlight implications for improving both the provision and use of online information. While many of these findings are generalizable to offline decisions, they are especially relevant to the technology-supported decisions that have become increasingly prevalent over the Internet.

Table of Contents

Introduction

Imagine, if you will, visiting a website such as TripAdvisor.com to check out the ratings that people have given different hotels before deciding where to stay at an upcoming conference in Memphis. Making the best of a bad destination (no offense to Elvis fans), you want to select a hotel that is clean and comfortable, and assume that the ratings of previous hotel guests will be more informative than claims on a hotel's own website! But how does the organization of information on the site affect its importance to you? How do you interpret the ratings provided by previous hotel guests, and do some opinions in the distribution stand out to you as more likely or impactful outcomes, thus deserving greater attention? Lastly, knowing that others have provided ratings conditional on their own experiences, does this change the consideration you should give to more extreme opinions? These are several of the questions examined here, with the broader objective of understanding the way that some information may be disproportionately salient to decision-makers, therefore receiving undue weight in their choices.

Even staunch rationalists would not deny that people often rely on simplifying heuristics to distill meaning from environmental cues and make actionable decisions, especially in situations characterized by abundant or conflicting information. In nearly every choice from selecting an entrée to selecting a mate, objectively dominant options are rare and it is taxing both cognitively and emotionally to reason ad nauseum about the required tradeoffs, so instead individuals convince themselves that one option is "best" based on some subset of salient criteria. Although this strategy sometimes generates suboptimal results, for example choosing a restaurant's "catch of the day" without noticing a tastier alternative, it is a parsimonious method for coping with large amounts of information in a manageable fashion. If one were to carefully weigh all factors of every available course of action – no matter how trivial the consequences – before deeming any to be satisfactory, he would certainly languish in indecision.

However, people in today's world experience a level of information overload orders of magnitude beyond that which ingrained responses have been evolutionarily adapted to tolerate. It is not merely the quantities of information that have grown exponentially in recent decades, but also the types of information that are available over the Internet. Never before have people had access to such extensive social networks that allow them to learn about the experiences of others, and furthermore to glean feedback from others on what those experiences were like. As a result, forums for opinion sharing offer an entirely new realm of information that can be taken into consideration by decision-makers. Additionally, the web permits comparison of options on more dimensions, with the display of this information often determined by the whims of designers. Individual biases may even be exacerbated in an online setting due to these inherent uncertainties, the study of which will hopefully shed light on the psychology of decision-making more generally.

The essays here explore several aspects of information aggregation that have grown in relevance alongside the Internet economy. The first paper looks at the role of interface design as another factor affecting the way that individuals integrate information in online contexts. A survey of existing websites first demonstrates the various – and seemingly arbitrary – ways that attributes are partitioned into discrete categories when displayed to consumers. A series of laboratory studies then validates that these groupings in fact play a role in the weight that decision-makers place on different option attributes, both in their explicitly reported importance and their implicit importance revealed through choices. The primary finding is that individuals place greater weight on attributes that receive a greater "share" of the categorization, perhaps because they anchor on an equal weighting across the set of available categories and insufficiently adjust weightings according to personal considerations. Compounding this effect, decision-makers may infer that a more knowledgeable designer is signaling the greater importance of certain attributes over others via their hierarchical grouping.

The second paper focuses on the choices people make based on one type of information available to them: opinions of others. Specifically, we ask how decision-makers take into account ratings variance. While high variance options – about which there is little consensus amongst raters – offer greater potential for a good outcome, they also pose greater risk of realizing a poor outcome. In fact, the attention that decision-makers pay to extreme possibilities appears to differ across choice domains: In generally positive domains (e.g., desserts), individuals prefer high variance options that maximize their chances of achieving the most favorable outcomes. On the other hand, in negative domains (e.g., disgusting foods), individuals prefer low variance options in avoidance of unfavorable outcomes. The proposed mechanism relates to the salience of these extremes: Because reference points are set by readily available memories of very good experiences in positive domains and very bad experiences in negative domains, most outcomes are seen as losses in positive domains (where utility is shown to be convex) and as gains in negative domains (where utility is concave). In addition, decision-makers believe that extreme outcomes are more probable for themselves than for the population of raters, suggesting that the observed preferences for variance may reflect a systematic bias in the weighting of opinion information.

Expanding on this theme, the third paper examines incentives for both the provision and use of opinion information, helping to explain its increasing influence in online communities. Since the motivation to share an opinion is conditional on the experience one is evaluating, the mere frequency of outcomes expressed does not necessarily mirror their rate of occurrence within the broader population of consumers. For this reason, the diagnosticity of opinion information is not obvious, yet it clearly helps to shape expectations and consumption choices. In turn, exposure to others' opinions will impact the perceived experiences and subsequent opinion sharing of those who do consume. From a theoretical standpoint, these causal effects lead to complex dynamics in the distribution of publicly available opinions and trends in the popularity of options they describe. In practice, individual consumers face the equally complex task of inferring meaning from opinions they encounter online.

Each of these papers highlights a unique challenge of information processing faced by members of online communities, and indeed the corresponding challenges faced by researchers of behavioral decision-making in understanding how choices are made on the Internet. Although fundamental questions about how people simplify the world around them remain relevant, the emerging capacity to continually alter information content and presentation, and even personalize it for different visitors to a website, adds new degrees of freedom to influencing consumer choice. The implications are profound: A more thorough understanding of how information is used will better equip its purveyors to make certain features more salient. This can occur either for the benefit of marketers who wish for particular information to be given greater attention, or for the benefit of market participants who currently bear most of the burden to filter and interpret the overwhelming wealth of information at their fingertips.

Chapter 1

Shaping Consumer Criteria Weighting and Choice through Attribute Partitioning

One of the many features that social networking sites such as Facebook and MySpace have added to the process of acquaintanceship is a feature whereby one can click on a preference of some user (for instance, that person's favorite television show from the 1980's), and receive a list of other users who share that same preference. In the real world, discovering fellow devotees of the *Golden Girls* is a difficult proposition; these search interfaces, on the other hand, offer instant access to such information. While the effectiveness of using television preferences to find friends remains to be seen, there is little doubt that such interfaces change not just the options most readily viewed, but more fundamentally, the weight accorded to different kinds of information: If a website encourages people to use television shows to find friends, we suggest, they are likely to use it, even if that attribute is not one that they have ever used – or wanted to use – before.

Imagine a social networking site that allows users to search on just four demographic attributes: age, height, weight and income, which users valued equally such that they would assign each attribute 25% of the total decision weight (see Figure 1.1a).

Figure 1.1a

Attribute	Decision Weight
Age	25%
Height	25%
Weight	25%
Income	25%

Adding "television preferences" to this list of four would, the research we review below suggests, cause participants to divert some of the weight accorded to the existing four demographic attribute to this new preference attribute; if they equally weighted each attribute again, this new attribute would receive 20%, and each of the initial four would "lose" 5% (see Figure 1.1b), meaning that those initial attributes would now be less influential in subsequent choices.

Figure 1.1b

Attribute	Decision Weight
Age	20%
Height	20%
Weight	20%
Income	20%
Television Preferences	20%

In addition, the inclusion of this new category of preference attributes might, encourage users to group these demographic attributes together – and of course websites themselves might encourage such groupings. In this case, we might expect the four demographic attributes to be afforded just 50% of the total (each receiving 12.5% on average), while the preference attribute might receive a full 50% (see Figure 1.1c). In this (admittedly extreme) case, this new preference attribute would dominate any decision, even though participants' prior valuation of this attribute may have been quite low.

Figure 1.1c

Attribute	Decision Weight
Demographics (Age, Height, Weight, Income)	50%
Television Preferences	50%

A quick glance at existing websites reveals just this kind of wide variability in the criteria emphasized in search interfaces. As just one example, consider several major online vendors of digital cameras: Amazon, Best Buy, Circuit City, and Ritz Camera. These sites each allow shoppers to filter their options according to preferences; as Figure 1.2 shows, however, the criteria differ widely from one site to the next, in both number (Circuit City allows search on three attributes, Amazon seven) and type (all four offer "brand", but Best Buy offers "color" but not "image stabilization," while Amazon offers the reverse). In addition, these websites easily could group several features under one heading, for example by placing "image stabilization" and "optical zoom" under the umbrella category "features." A consumer looking for a camera might engage in very different search processes depending on which website she used to perform that search – and a great deal of research reviewed below suggests that she might therefore end up with a very different choice set merely because of that different search process, even if all websites offered the same selection of cameras.

Figure 1.2

	Amazon	Best Buy	Circuit City	Ritz Camera
Brand				
Camera Needs				
Color				
Current Offers				
Display Size				
Features				
Image Stabilization				
Mega-Pixels				
Optical Zoom				
Price				
Status				
Viewfinder Type				

Shading indicates the presence of search criteria on different vendor websites (as of February 2008).

Thus by subtle alterations in how attributes are grouped, websites can shift consumers' attention to different attributes. In this paper we explore how the mere presentation of attributes – holding the attributes themselves constant – can impact their importance to consumers. We test this hypothesis not simply by adding new attributes to existing attributes (as in Figure 1.1b), but using a design in which we hold attributes constant and alter consumers' perceptions of how much weight they should be accorded (as in Figure 1.1c). In short, we explore not just how adding the ability to search for friends via their television preferences alters search behavior, but rather how adding that attribute changes consumers' *underlying valuation* of the attribute "television preferences” in selecting friends, altered valuations which we show guide subsequent choice.

Context-Dependent Choice

While we focus on a case in which websites can predictably reshape consumer preferences, not all presentations of search criteria lead people down the wrong path: in a world of abundant product choices, organizing information can serve the very useful purpose of helping people filter out irrelevant attributes and hone in on those attributes about which they care most. Indeed, many online retailers design interfaces in a deliberate effort to improve the search process for consumers (Bakos, 1997; Hearst, 2006; Schafer, Konstan, & Riedl, 1999). Grouping options can assist consumers in part because categories provide important information about the shared attributes of items in that category (Huber & Kline, 1991; Roberts & Lattin, 1991), which can then help choosers refine their set of options (Chakravarti & Janiszewski, 2003; Diehl, 2005; Diehl & Zauberman, 2005; Zhang & Fitzsimons, 1999); in Diehl, Kornish, and Lynch (2003), for example, providing users with screening devices improved their choices.

At the same time, however, changes in how information is presented can also change consumer preferences, in ways that may not always lead to better outcomes. A large body of research has explored the tendency for individuals to construct their preferences based on whatever information happens to be salient in the environment (Ariely & Norton, 2008; Bettman, Luce, & Payne, 1998; Payne, Bettman, & Johnson, 1992), such as inferring the attractiveness of options from contextual cues about what is available (Prelec, Wernerfelt, & Zettelmeyer, 1997). Subtle changes in online interfaces frequently bring some options or attributes to the forefront (Kleinmuntz & Schkade, 1993; Lurie & Mason, 2007); Mandel and Johnson (2002), for example, showed that visual priming on websites makes some information more focal to decision-makers. Indeed, Tversky's (1972) seminal work on Elimination by Aspects has at its core the notion that the order in which information is considered can determine the option that is ultimately chosen (Bettman & Kakkar, 1977; Chakravarti, Janiszewski, & Ulkumen, 2006). Online search interfaces of necessity require web designers to make a number of decisions about which attributes to include or exclude (should television show preferences be included?) and how salient to make such attributes (should they be presented alone or grouped with other preferences such as music and movies?): Given research on the impact of such environmental cues on the malleability of preferences, there is little doubt that such decisions impact consumer choice (see Johnson, Moe, Fader, Bellman, Lohse, 2004).

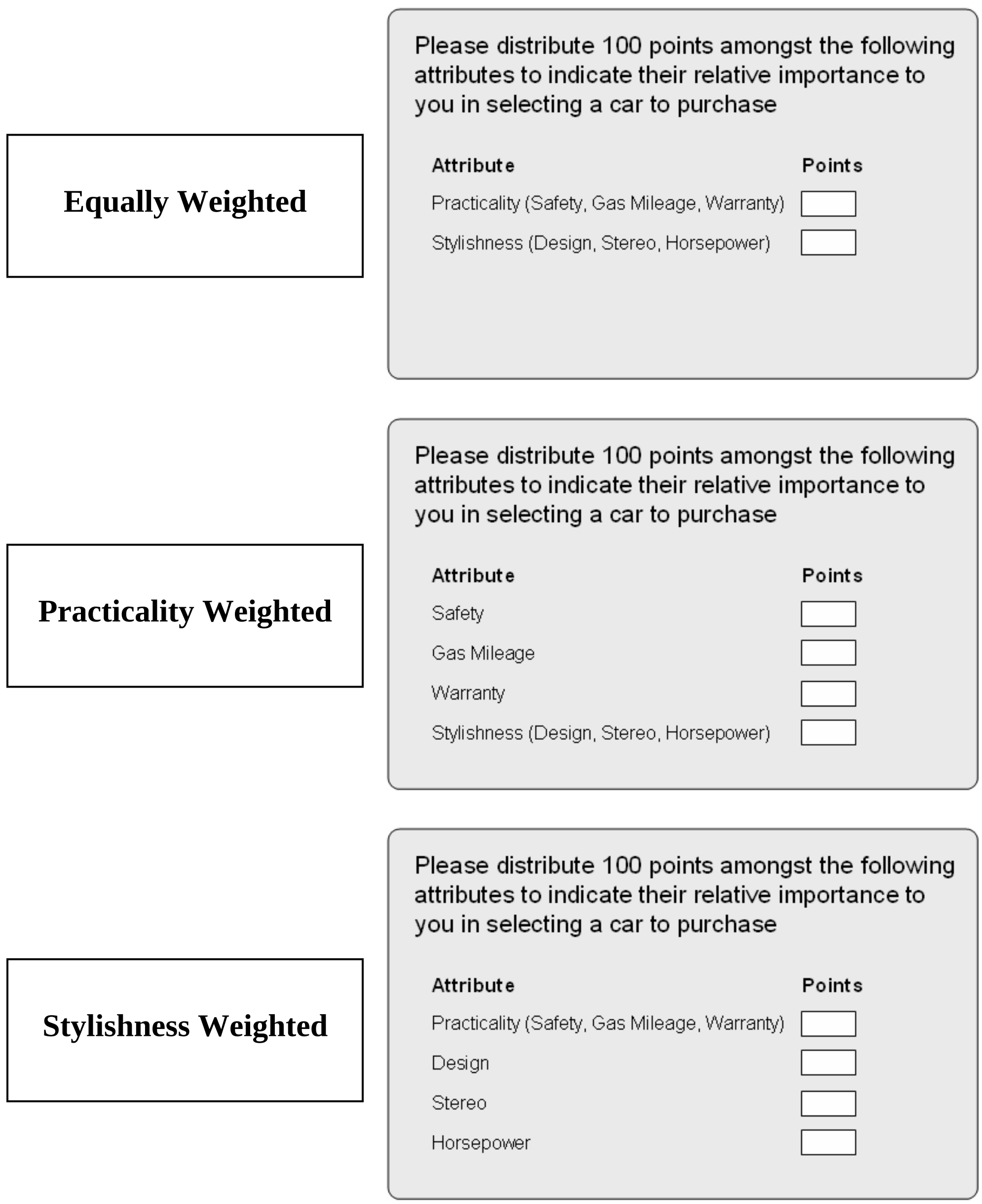
The Current Research

In this paper, we extend previous research exploring how the presentation of options influences choice by demonstrating the role that presentation can play in reshaping the *inputs* to choice, impacting the value that people place on the attributes that underlie their choices. First, a growing body of research has documented a "diversification bias," the tendency for individuals to spread their attention – and consumption – evenly across available sets of options (Read & Loewenstein, 1995; Simonson, 1990; Simonson & Winer, 1992). For instance, in an experiment in which participants chose between five investment funds, participants presented with four equity funds and one fixed-income fund allocated 68% to equities, while those presented with just one equity fund and four fixed-income funds allocated just 43% to equities (Benartzi & Thaler, 2001). Relatedly, research on partition dependence shows that partitioning information even more explicitly into different categories – such as grouping wines by grape compared with region – can have marked influences on subsequent choices, with consumers diversifying more across different grapes when wines are grouped by grape, but diversifying more across different regions when wines are grouped by region (Fox, Ratner, & Lieb, 2005).

Again, we propose – and the studies we present below demonstrate – that diversification and partition impact not only choices between options, but the *inputs* to those choices. We test this hypothesis using a paradigm in which we hold attributes constant and simply alter consumers' perceptions of how much weight they should be accorded, and then examine the impact of such these altered attribute valuations on choice. Figure 1.3 demonstrates our basic paradigm for someone interested in buying a new car: Note that in each version, the same attributes are present, but are given different weight. We predict that in the "Practicality Weighted" version, the three practicality traits presented separately will be accorded greater total weight in a decision than when they are all grouped under the "Practicality" category – as in the "Stylishness Weighted" version – and that therefore the extent to which subsequent choices are made on the basis of practicality will be greater in the former than in the latter version.

The first two studies explore how different attribute partitions change the value participants place on them, for choosing both cars and people to date (Study 1). Study 2 then demonstrates the impact of these partitions on choices between hotels, exploring whether these changed valuations influence decisions. Finally, Study 3 explores whether consumers may be dissatisfied with their choices when influenced by partitions – an important consideration for marketers – as well as showing how our paradigm can be integrated with existing online recommendation agents.

Figure 1.3



Example of three attribute presentations (used for “Choosing a Car” in Study 1).

STUDY 1

In this first study, we wanted to establish our basic effect, that partitioning attributes can change the weight that people place on them. We expected that partitioning attributes would change the total value accorded to such traits, in comparison to a condition which did not differentially weight the two categories.

We explored this phenomenon in two domains: Choosing a car to buy, and, in a domain more familiar to our primarily college-aged students, choosing someone to date (Frost, Chance, Norton, & Ariely, in press; Norton, Frost, & Ariely, 2007).

Method

Participants (N = 98, 52 female, M_{age} = 22.7) received $20 to complete this study along with several unrelated studies.

Choosing a Car. Participants were first asked to imagine they were considering the purchase of a new car, and to distribute 100 points across various attributes to indicate their relative importance in making that decision. Each participant was randomized into one of three conditions: those in the "Equally Weighted" condition saw the three practicality attributes grouped into a single category and the three stylishness attributes grouped into another category, those in the "Practicality Weighted" condition saw each practicality attribute listed separately but all stylishness attributes grouped, and those in the “Stylishness Weighted” condition saw each stylishness attribute listed separately but all practicality attributes grouped (see Figure 1.3).

Choosing a Date. Next, participants were asked to imagine they were considering choosing someone to date. Each participant was randomized into one of three conditions: those in the “Equally Weighted” condition saw three personality attributes grouped into a single category and three appearance attributes grouped into another category, those in the “Personality Weighted” condition saw each personality attribute listed separately but all appearance attributes grouped, and those in the “Appearance Weighted” condition saw each appearance attribute listed separately but all personality attributes grouped.

Results and Discussion

Choosing a Car. In order to compare the relative importance ascribed to a car’s practicality versus stylishness, we computed the sum of points given to practicality attributes in all versions (or, equivalently, 100 minus the sum of points given to stylishness attributes). As expected, weighting of the two types of information varied by condition, $F(2, 95) = 13.1$, $p < .0001$ (see Table 1.1). In the “Equally Weighted” condition, participants demonstrated a slight preference for practicality over stylishness ($M = 68.3$, $SD = 21.0$); as predicted, however, when practicality attributes were broken out into separate categories but stylishness attributes remained grouped in the “Practicality Weighted” condition, this preference for practicality increased ($M = 73.7$, $SD = 22.2$), but when stylishness attributes were broken out into separate categories and practicality remained grouped in the “Stylishness Weighted” condition, the preference for practicality was reduced (*M* = 47.4, *SD* = 23.5), to such an extent that participants now gave more overall weight to stylishness.

Table 1.1: Average number of points given to attributes when "Choosing a Car" (Study 1).

		Equally Weighted	Practicality Weighted	Stylishness Weighted
Practicality		68.3		47.4
	Safety		22.0	
	Gas Mileage		35.7	
	Warranty		16.0	
Stylishness		31.7	26.3	
	Design			25.9
	Stereo			13.6
	Horsepower			13.1
	Dependent Variable (Sum of Practicality Points)	68.3	73.7	47.4

Choosing a Date. We again created a metric by computing the sum of points given to personality attributes in all versions). Once again, the relative weighting of personality differed across conditions, $F(2, 95) = 30.2$, $p < .001$ (see Table 1.2). Participants in the "Equally Weighted" condition were somewhat more concerned about personality than appearance (*M* = 58.6, *SD* = 15.9). For those in the "Personality Weighted" condition, however, this preference was even stronger (*M* = 75.7, *SD* = 13.1), while those in the "Appearance Weighted" actually viewed personality as *less* important than appearance (*M* = 46.5, *SD* = 17.6).

Table 1.2: Average number of points given to attributes when "Choosing a Date" (Study 1).

		Equally Weighted	Personality Weighted	Appearance Weighted
Personality		58.6		46.4
	Intelligence		30.5	
	Sense of Humor		21.6	
	Kindness		23.6	
Appearance		41.4	24.3	
	Body			23.3
	Face			21.9
	Hair			8.4
	Dependent Variable (Sum of Personality Points)	58.6	75.7	46.4

STUDY 2

Both versions of Study 1 – choosing cars or dates – provide support for our hypothesis that decision-makers can be swayed away from their preexisting attribute preferences, weighting attributes differently depending on their presentation. In Study 2, we explore whether our paradigm can also influence choices between options, as our account predicts. In most real-world choices, of course, individuals make their selections without explicitly stating the relative importance they place on different attributes (as in the first two studies). In Study 2, therefore, we presented participants with a more realistic task – choosing a hotel – in which we altered the presentation of attribute information, but gave them only the task of choosing the hotel they preferred, as they would be likely to do if visiting an actual website.

Method

Participants (N = 124; 69 female; M_{age} = 24.0) received $20 to complete this computer study along with several unrelated studies. They were randomized into three conditions: The "Equally Weighted" condition aggregated both room attributes (cleanliness and comfort) and hotel attributes (service and condition), the "Room Weighted" condition grouped the hotel attributes but presented the two room attributes separately, and the "Hotel Weighted" condition grouped the room attributes but presented the two hotel attributes separately.

Participants were asked to choose one of ten hotel options based on ratings for each of the categories displayed. Across all conditions, ratings for the each of the four attributes – cleanliness, comfort, service, and condition – of the ten options were generated randomly on a 5-point scale. We counterbalanced whether room or hotel attributes appeared on the left or right, which did not impact the analyses below so we do not report it further; in addition, the ten options were presented in random order.

We structured these ten options such that five options were always stronger on room attributes (uniform ratings between 3 and 5 for both room cleanliness and comfort) but weaker on hotel attributes (uniform ratings between 1 and 3 for both hotel service and condition); this was reversed for the other five options, which were therefore stronger on hotel attributes and weaker on room attributes. For conditions which aggregated two attributes, the ratings of each attribute were averaged to compute the category rating. Note that since each condition contained the same underlying information, participants had the option to "unpack" the ratings which were aggregated into a single category; by noticing that hotel service and comfort were grouped, for example, a participant could have weighted the aggregated rating twice in her decision.

Table 1.3 contains an example of the options that might have been offered to a participant. In this example, Hotels 1, 3, 4, 5, and 9 are the random five hotels that are stronger on room attributes than overall hotel attributes; we would predict that participants in the "Room Weighted" condition would be more likely to pick one of these options than participants in the "Hotel Weighted" condition. After reviewing ratings, all participants indicated which option they would choose.

Table 1.3: An example of the hotel options shown to participants in Study 2.

	Room Cleanliness	Room Comfort	Hotel Service and Condition
Hotel 1	4.1	4.7	1.8
Hotel 2	1.8	2.1	4.4
Hotel 3	4.8	3.1	2.2
Hotel 4	3.8	4.2	1.8
Hotel 5	3.3	3.2	1.4
Hotel 6	2.6	1.1	3.9
Hotel 7	1.6	2.3	3.7
Hotel 8	1.8	2.6	4.5
Hotel 9	4.3	4.2	2.4
Hotel 10	1.9	2.5	3.9

Results and Discussion

We computed a binary dependent variable which indicated whether a participant selected a hotel that was stronger on room attributes. As we expected, participants' hotel selections differed across conditions, $X^2(2, 124) = 6.58$, $p < .04$. In the "Equally Weighted" condition, participants exhibited a strong preference for options stronger on room attributes (84% of choices); as predicted, this preference for hotels with better rooms increased even further (98%) in the “Room Weighted” version, but decreased (80%) in the “Hotel Weighted” condition.

STUDY 3

Study 2 demonstrated that the redistribution of attribute weights we observed in Study 1 appears to influence choice; despite a strong preference for better rooms over other hotel attributes, these preferences were still impacted by the way in which attributes were partitioned. In Study 3, we combine the paradigms from the first two studies, asking participants both to explicitly weight the importance of attributes and to make a subsequent choice based on these attributes. As in Study 1, we use a domain both familiar and relevant to our participants: choosing someone to date.

In addition, we explore the practical implications for marketers, by examining whether the shifts in attribute weights caused by different attribute partitions can be integrated with recommendation agents, such that when participants provide attribute weights, recommendation agents can then use those weights to present options that accommodate those altered preferences. Indeed, decision support interfaces are frequently tailored to individuals’ unique interests based on observable prior behavior (Adomavicius & Tuzhilin, 2005; Ansari, Essegaier, & Kohli, 2000; Hirsh, Basu, & Davison, 2000; Spiekerman & Paraschiv, 2002). Given that our experimental manipulations *alter* people’s prior decision weights, however, we wanted to ensure that participants did not feel duped or prodded into making suboptimal decisions. Because trust in the integrity of such agents is crucial (Fitzsimons & Lehmann, 2004; Pu & Chen, 2006; Smith, Menon, & Sivakumar, 2005), and because even very subtle changes in categorization can impact consumer satisfaction (Mogilner, Rudnick, & Iyengar, in press), we also assessed participants' satisfaction with their choices.

Method

Participants (N = 76; 36 female; M_{age} = 22.1) received $20 to complete this computer study along with several unrelated studies. As in Study 1, they were instructed to distribute 100 points across attributes to indicate their importance in choosing someone to date, and were assigned to the "Equally Weighted," "Personality Weighted," or "Appearance Weighted" conditions; we used the same six attributes as in Study 1.

Unlike Study 1, however, participants next viewed three possible dating options from which to choose. Each option had ratings for all six attributes on 10-point scales (see Table 1.4 for the three options). We calculated the expected value of each option based on each participants' distribution of 100 points, and then presented the options in order of expected value from highest to lowest.[11] Thus the three options were always the same for all participants, but were merely shown in different orders depending on participants' previous point distributions.

[1] For example, if a participant in the "Personality Weighted" condition allotted 25 points to each of the four available categories (intelligence, sense of humor, kindness, and aggregated appearance attributes), a date option with a rating of 5 on each attribute would have an expected value of 0.25*5 + 0.25*5 + 0.25*5 + (0.25/3)*5 + (0.25/3)*5 + (0.25/3)*5 = 5. We assumed an equal distribution of points across component attributes when points were only given to the grouped category.

Table 1.4: The three dating options shown to participants in Study 3; options were presented according to participants' attribute ratings.

	Intelligence	Sense of Humor	Kindness	Body	Face	Hair
Person 1	5	5	5	6	5	9
Person 2	8	3	4	7	6	5
Person 3	4	5	5	9	7	4

After selecting one of the three options, participants were asked on the final screen to report satisfaction with their choice on a 7-point scale (1: *very dissatisfied* to 7: *very satisfied*).

Results and Discussion

Our first dependent measure was the sum of points given to personality attributes. As in Study 1, the weighting of personality attributes varied across conditions, $F(2, 73) = 42.2$, $p < .001$. In the "Equally Weighted" condition, there was no preference for personality over demographic information ($M = 51.0$, $SD = 15.1$), but preference for personality traits increased in the "Personality Weighted" condition ($M = 75.5$, $SD = 12.0$), and decreased significantly in the "Appearance Weighted" condition ($M = 35.1$, $SD = 18.9$).

In addition, these differences in point allocations impacted choice between date options. Preferences for the three options varied as a function of condition, $X^2(4, 76) = 9.48$, $p = .05$, because participants tended to select the option that we had "designed" for them to find most appealing based on their distribution of points. Person 1 was the most popular selection (chosen by 48% of participants) in the "Equally Weighted" condition, Person 2 was the most popular choice (36%) in the "Personality Weighted" condition, and Person 3 was the most popular choice (50%) in the “Appearance Weighted” condition.

Finally, we explored whether our participants might be unhappy about having been duped, reacting against us presenting options according to their (experimentally-manipulated) point distributions. This was not the case, as participants were equally satisfied with their selections across all three conditions, with means ranging from 4.36 to 4.46, $F(2, 73) = .06$, *ns*.

GENERAL DISCUSSION

The above studies illustrate the powerful impact that organization of information can have on choice. Specifically, our studies show that the attention that different attributes receive depends on how they are partitioned: Attributes that are displayed as separate categories tend to receive greater weight, whereas those that are grouped together under umbrella categories are discounted as less important. Study 1 demonstrates this phenomenon when individuals are asked to explicitly distribute points across attribute categories, while Study 2 demonstrates the impact of such groupings on choice. Finally, Study 3 shows how our effects can be combined with recommendation agents to first impact underlying valuations of attributes and then provide consumers with options that cater to (and reinforce) those altered preferences. We thus demonstrate that diversification bias and partition dependence – previously shown to affect distribution of consumption across options – can also be extended to impact people's distribution of decision weights across the attributes that shape those choices.

Given the ease with which websites can change the presentation of attributes, these findings have immediate implications for the design of retail websites, though we hasten to add that these changes in design can be either consumer-focused or retailer-focused. For instance, a web designer with good knowledge of which attributes are most important to ultimate consumer satisfaction could present attributes in a way that signals their appropriate relative weighting, thus enabling users to make decisions in their own best interests (Benbasat & Todd, 1992; Todd & Benbasat, 1994). On the other hand, a designer who wishes to influence choice for his own profit (e.g., to sell over-stocked items) can drive users to weight attributes such that they select suboptimal products; our results from Study 3 – in which consumers were equally satisfied with their choices regardless of how attributes were grouped – suggest that consumers might be relatively insensitive to such manipulations. Indeed, if individuals distort the weighting of various attributes according to how they are grouped, they may also be likely to use the same weighting to assess the quality of outcomes, meaning that their expected and actual utility might correspond closely.

Further Opportunities

We have focused primarily on one case in which search interfaces impact choice, but in some sense the possibility for such interfaces to change decision-making are limitless – as are the opportunities for experimentation to explore these possibilities. For example, while we demonstrate how changing the grouping of multiple attributes changes the valuations of those attributes, even modifying single attributes in isolation can change consumer search processes. Consider several major online vendors of digital cameras: Amazon, Best Buy, Circuit City, and Ritz Camera. All four sites allow shoppers to filter their options by price, but the exact price ranges for inclusion/exclusion vary, and thus could lead to different choice sets even if this was the only attribute used. Figure 1.4 shows the price ranges for each of the four sites in September 2007. A consumer who searches for cameras that cost roughly $300, for example, would be left with cameras ranging from $200-$499 at Amazon, but a much narrower range of $300-$399 at Ritz Camera.

Figure 1.4

<table>
<tr><th></th><th>Amazon</th><th>Best Buy</th><th>Circuit City</th><th>Ritz Camera</th></tr>
<tr><td>$0-24</td><td></td><td rowspan="2"></td><td rowspan="5"></td><td rowspan="3"></td></tr>
<tr><td>$25-49</td><td></td></tr>
<tr><td>$50-99</td><td></td><td></td></tr>
<tr><td>$100-149</td><td rowspan="2"></td><td></td><td rowspan="2"></td></tr>
<tr><td>$150-199</td><td></td></tr>
<tr><td>$200-249</td><td rowspan="4"></td><td></td><td rowspan="2"></td><td rowspan="2"></td></tr>
<tr><td>$250-299</td><td rowspan="3"></td></tr>
<tr><td>$300-399</td><td rowspan="2"></td><td></td></tr>
<tr><td>$400-499</td><td></td></tr>
<tr><td>$500-749</td><td rowspan="2"></td><td></td><td rowspan="10"></td><td></td></tr>
<tr><td>$750-999</td><td></td><td></td></tr>
<tr><td>$1000-1249</td><td rowspan="3"></td><td></td><td></td></tr>
<tr><td>$1250-1499</td><td></td><td></td></tr>
<tr><td>$1500-1999</td><td></td><td></td></tr>
<tr><td>$2000-2499</td><td rowspan="3"></td><td></td><td rowspan="5"></td></tr>
<tr><td>$2500-2999</td><td></td></tr>
<tr><td>$3000-4999</td><td rowspan="3"></td></tr>
<tr><td>$5000-9999</td><td></td></tr>
<tr><td>$10000+</td><td></td></tr>
</table>

Horizontal bars indicate dividers between price ranges for online vendors of digital cameras. Shading indicates the range of prices that would be included for a consumer searching for a $300 camera.

But do such groupings impact decision-making? We asked a new set of participants (N =164) to imagine they were looking for a date, and to indicate the marital status that they considered acceptable in a potential partner. When we presented participants with just two options – "never married" or "married in the past" – just 17% included potential dates who were "married in the past," even though they were allowed to check all acceptable categories. When we presented a different set of participants with the "never married" option as before, but replaced the "married in the past" option with three additional options – "currently separated," "widowed," and "divorced" – some 38% included at least one of the three, which, of course, are subsets of the "never married" option in the first version. This change in share from 17% to 38% – based solely on how this attribute was broken out – suggests the potential for such subtle changes to impact choice, and offers a promising direction for future research.

Chapter 2

Variance-Seeking for Positive (and Variance-Aversion for Negative) Experiences

Individuals faced with some decision (for example, which movie to see on a Friday night) frequently assess the opinions of others to help them decide; in its most basic form, this process involves simply counting the number of people who like each option (the number of people in line for each movie). While the notion that people's preferences and beliefs are influenced by the preferences and beliefs of others is not a new one (Cialdini and Goldstein, 2004; Marsden and Friedkin, 1993), the sheer number of available opinions has increased exponentially in recent years, with consumers' instant online access to the seemingly limitless – and often conflicting – views of strangers (Godes and Mayzlin, 2004; Salganik, Dodds, and Watts, 2006). This research examines an understudied aspect of the impact of social information on preferences, one which has become increasingly relevant with the rapid rise of Internet ratings for many types of products and services, from movies and music to hotels and travel destinations: How consumers incorporate *variance* in opinions – endemic to any mass of continuous ratings – into their own preferences.

If all consumers choose one movie over another, the obvious inference is that the first movie is superior. But what about the common case in which similar numbers of people choose both movies, and provide roughly similar ratings on average, yet the variance of those ratings differs? While others' preferences are often observed as discrete choices, online ratings provide detailed information about the *distribution* of those ratings. Two movies may have the same 3-star mean rating, but while Movie A has very low variance (all viewers give it 3 stars), Movie B has very high variance (with ratings ranging from 1 to 5 stars); choosing a movie with high variance carries both risk and reward, while choosing the low variance option offers a surer bet. Knowing whether and when consumers are variance-seeking and variance-averse has clear implications for marketers trying to predict consumer behavior. Under what conditions do consumers prefer high-variance options, and when might they prefer low-variance options? More generally, how does uncertainty induced by the variability of others' opinions impact consumers' preferences?

Risk-Seeking in the Domain of Gains?

In line with much previous research, we suggest that people's preferences for variance in ratings will be influenced by the valence of the decision domain. In particular, we propose that people demonstrate a general variance-aversion in negative domains, playing it safe when choosing negative experiences, but variance-seeking in positive domains, choosing to gamble when choosing positive experiences. Students of Prospect Theory (Kahneman and Tversky, 1979) may sense a tension with one of the central tenets of that theory, that people are generally risk-seeking in the domain of losses, but risk-averse in the domain of gains. Indeed, many researchers have conflated "losses" with "negative experiences" – citing evidence that "bad is stronger than good" (Baumeister, Bratslavsky, Finkenauer, and Vohs, 2001; Peeters and Czapinski, 1990) simultaneously with Kahneman and Tversky (1979). On the contrary, we suggest that decision-makers may actually respond to negative experiences as gains and positive experiences as losses: While losing money in gambles is negative, and winning money in gambles is positive, this does not mean that all negative experiences are viewed as potential losses and all positive experiences as potential gains. In fact, as we will outline below, we suggest that *most* negative experiences may be viewed as potential gains, and *most* positive experiences as potential losses.

Understanding the seeming contradiction between "gains and losses" and "positive and negative experiences" eliciting different preferences for risk involves a deeper consideration of the other central contribution of Prospect Theory – that gains and losses are determined by psychological reference points, rather than by the valence of some domain. Outcomes are generally treated as losses whenever they fall below some reference point, but gains when they exceed that reference point (March and Shapira, 1992; Payne, Laughhunn and Crum, 1980; West and Broniarczyk, 1998), with implications for preferences for risk. In Budescu, Kuhn, Kramer, and Johnson (2002), people displayed vagueness-seeking for positive gambles and vagueness-aversion for negative gambles because they focused on attaining the very best outcomes in positive domains and avoiding the very worst outcomes in negative domains (see also Bettman, 1973).

But why might reference points be set so high in positive domains, and so low in negative domains? The fact that people focus on extreme reference points in Budescu et al. (2002), while explaining people's risk preferences, does not answer the question of why reference points are so extreme in the first place. After all, people have a wide range of experiences, and so we might expect that their reference points lie somewhere around their mean experience across all of their outcomes. Returning to our opening example, people have likely seen a wide range of movies – poor, average, and good – yet we suggest that they set their expectations at the high end of the spectrum, which then drives their risk-seeking behavior. West and Broniarczyk (1998) for example, showed that consumers preferred movies about which critics disagreed to ones about which there was consensus; because consumers' reference points were high, only when there was variance in critic ratings were ratings high enough to exceed consumers' reference points.

We suggest that biases in how people recall past experiences – and the resultant impact on their utility functions – account for their tendencies to set reference points very high in positive and very low in negative domains, leading to variance-seeking and variance-aversion, respectively. Certainly, reference points are influenced by memories of one's previous experiences (Stewart, Chater, and Brown, 2006), which then influence subsequent decisions (Koszegi and Rabin, 2006; Novemsky and Dhar, 2005; Thaler and Johnson, 1990). Importantly for our account, people display consistent biases in their recollection of valenced experiences, remembering the most extreme of both kinds of experiences. When asked to recall past experiences, people recall not a random sample of experiences, but the very best Red Sox game they ever saw, or the most disastrous time they missed their train; in addition, people view these extreme experiences as typical suggesting that they are very likely to be used as reference points despite their actual rarity and extremity (Morewedge, Gilbert, and Wilson, 2005).

What are the consequences of this memory bias for extreme bad experiences in negative and extremely positive experiences in positive domains? We suggest that this over-reliance on extreme past instances to set reference points when choosing future experiences – recalling “1”s on a 10-point scale when thinking about negative experiences, but perfect “10”s on that scale when reflecting on positive experiences – shape people’s utility functions, and ultimately their preferences for variance. We expect utility curves to reflect people’s emphasis on extreme outcomes at both ends of the spectrum. Because people’s reference points for positive experiences are at the very high end of the scale, then movement at the top of that scale should offer more utility than movement elsewhere in the scale (e.g., moving from a “9” to a “10” creates more utility than moving from a “5” to a “6”), and thus utility is likely convex in positive domains: This utility function should therefore lead to variance-seeking in positive domains, as people seek that additional utility. In negative domains, in contrast, where reference points are at the very low end of the scale, movement at that low end has larger implications for utility (e.g., moving from a “2” to a “1” is much worse than moving from a “6” to a “5”), and thus utility is likely to be concave in negative domains; this utility function should lead to variance-aversion in negative domains, as people seek to avoid the potential for large decreases in utility at the low end of the scale. This account – which we test experimentally below – suggests that, in contrast to some previous conceptualizations, most negative experiences may be treated as gains, while most positive experiences may be treated as losses.

Shifting Reference Points with Regulatory Focus

If our account is correct, that reliance on memories of extreme experiences induces different reference points in positive and negative domains which subsequently drive variance-seeking and variance-aversion, then shifting those reference points away from those extremes should also impact preferences for variance, making people less likely to seek variance in positive domains and less likely to avoid variance in negative domains. In short, reducing disproportionate focus on extremely high ratings in positive domains and extremely low ratings in negative domains would lead to more equal weighting of the entire spectrum of possible outcomes, thus shifting reference points toward the distribution mean and mitigating preferences for variance. Altering people's regulatory focus offers one such approach (Higgins, 1997): While promotion-focused individuals tend to process information in an abstract fashion, those in a prevention focus tend to engage in more concrete processing (Forster and Higgins, 2005), precisely the kind of processing that would encourage deliberate consideration of all ratings – including those that do not fall at initially salient extremes – rather than a reliance on extreme memories. In two studies below, we both manipulate regulatory focus and use natural shifts in regulatory focus over the life span – with older individuals shifting toward a prevention focus later in life (Heckhausen, 1997; Lockwood, Chasteen, and Wang, 2005) – to explore the impact of this moderator.

Overview of Studies

We first show that utility is convex for positive experiences, consistent with reference points defined by extremely positive experiences, but conversely that utility is concave for negative experiences, reflecting reference points at the opposite endpoint of the spectrum (Study 1). Study 2 illustrates the impact of these utility functions on choice, as people exhibit variance-seeking in positive domains and variance-aversion in negative domains. In Study 3, we address an alternative account for our results, demonstrating that preferences for uncertainty are not driven by outlier ratings; variance-seeking in positive and variance-aversion in negative experiences persists for experiences with different variances but the same range of possible outcomes. Finally, Studies 4 and 5 examine the dampening of responses to variance for prevention-focused and older individuals.

STUDY 1: UTILITY CURVES FOR POSITIVE AND NEGATIVE EXPERIENCES

Our hypothesis for the reversal of variance preferences across domains was derived from previous research demonstrating that very positive and very negative extremes may be most salient (Morewedge et al. 2005), leading people to adopt very high reference points in positive domains and very low reference points in negative domains. In this study, we explore the shape of utility curves that result from such extreme reference points; in the remaining studies, we then demonstrate choice behavior in line with the behavior suggested by these utility curves. As we suggested above, if people do set reference points at extremes, then any experience below the top of the rating scale would be treated as a loss in positive domains, whereas any outcome above the bottom of the scale would be treated as a gain in negative domains, making utility convex in positive and concave in negative domains. In Study 1, we ask participants to construct their utility curves in positive and negative domains to examine whether these curves support our account.

Method

Participants (N = 147; 77 female, M_{age} = 23.1) completed this survey in the laboratory as part of a series of unrelated surveys. Participants were asked to imagine each of two scenarios, one in which they visited a bakery to purchase a dessert and another in which they were a contestant on the television game show Fear Factor. In response to each scenario, participants reported the amount they would pay to eat each of ten desserts (willingness to pay, WTP) and amount they would need to be paid to eat each of ten Fear Factor foods (willingness to accept, WTA) if they knew that they would rate the food in question each whole number from 1 = *very worst* to 10 = *very best*. We counterbalanced which category participants completed first (desserts or Fear Factor); in addition, participants were randomly assigned to report their dollar values for each food type starting at the worst outcome or starting at the best outcome. These variables did not impact our results and we do not report them further.

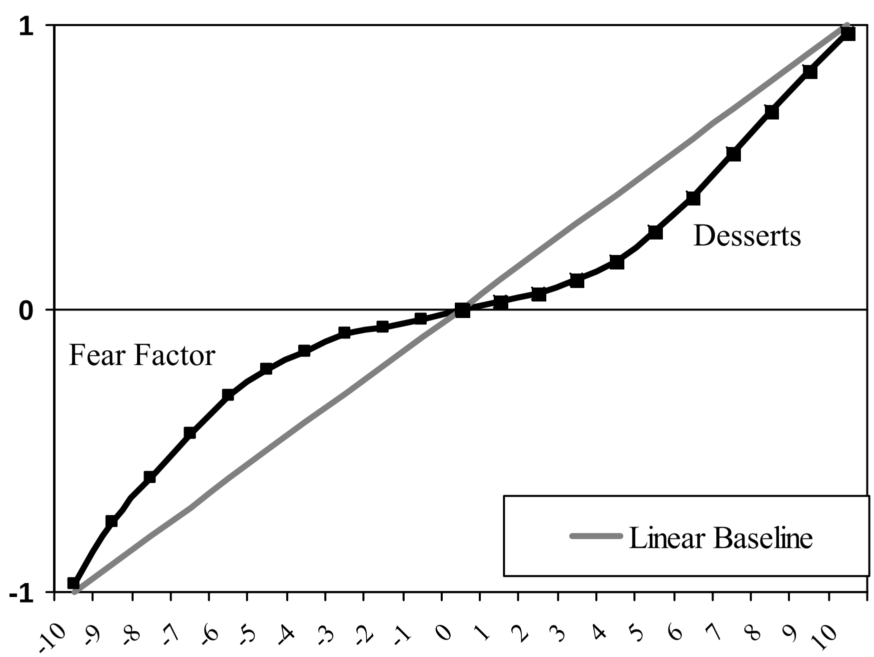
Due to large individual differences between participants in the magnitude of their WTP and WTA, we rescaled each participant's lowest response to 0 and highest response to 1 (with proportional ratings for intermediate responses) to standardize incremental changes corresponding to each rating additional rating level across all participants.

Results and Discussion

We fit a regression model to predict utility for desserts using willingness to pay as a proxy WTP = d0 + d1*$rating$ + d2*$rating^2$ + error, and a regression model to predict utility for Fear Factor foods using the negative of willingness to accept as a proxy -WTA

= d3 + d4**rating* + d5**rating*^{2} + error, including both rating and rating-squared terms as independent variables in each. Not surprisingly, the first term was positive in the both models (d1 = .021, $p < .01$; d4 = .23, $p < .001$); individuals expected greater utility from consuming higher rated foods. Most importantly, the squared term was positive in the model for desserts (d2 = .0082, $p < .001$), but negative in the model for Fear Factor foods (d5 = -.012, $p < .001$). In other words, utility outcomes indeed seems to be convex in the positive domain and concave in the negative domain, as shown in Figure 2.1. These curves are particularly notable in their seeming contradiction of standard utility curves across gains and losses for gambles; again, we suggest that this seeming contradiction is due to the fact that most positive experiences are actually treated as losses, and most negative experiences as gains.

Figure 2.1



Utility curves constructed by participants in positive (desserts) and negative (Fear Factor foods) domains, compared to a linear baseline (Study 1). For ease of presentation Fear Factor outcomes are graphed on the negative scale (-10 to -1).

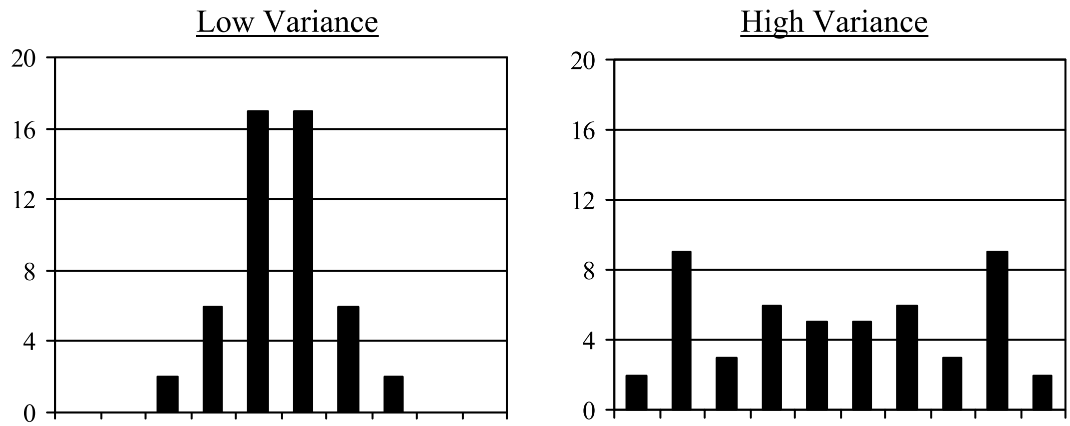
STUDY 2: VARIANCE-SEEKING IN POSITIVE AND VARIANCE-AVERSION IN NEGATIVE DOMAINS

In Study 2, we explore whether these different utility functions – in which attaining highly favorable outcomes in positive domains but avoiding highly negative experiences in negative domains have the greatest relative impact on utility – as a result lead to variance-seeking behavior in positive domains and variance-aversion in negative domains. We provided participants with one high and one low variance distribution for two foods, keeping the mean constant. We predicted that when these foods were labeled desserts (a positive domain), participants would prefer higher variance distributions; in contrast, we expected that labeling the two foods as disgusting (from the television show "Fear Factor") would induce a preference for low variance distributions.

Method

Participants (N = 113; 76 female, M_{age} = 34.0) completed this survey online as part of a larger set of unrelated surveys. We showed participants distributions of two foods, each containing the ratings of 50 individuals who purportedly had sampled these foods. Both distributions had the same average rating (M = 5.5) on a 10-point scale, but one was high variance (SD = 2.7) and one was low variance (SD = 1.1). In one condition, participants were told the two foods were "Fear Factor" foods, while in the other they were told both foods were desserts. We counterbalanced whether the high or low variance option appeared first (see Figure 2.2 for the distributions viewed by participants). Participants indicated which food they would rather eat if asked to choose.

Figure 2.2



Low and high variance distributions as shown to participants in Studies 2, 4 and 5. Bars indicate the number of previous consumers who gave that food the indicated rating. Distributions were labeled either "Fear Factor Foods" or "Desserts."

Results and Discussion

As predicted, while participants preferred the low variance option (64%) to the high variance option (36%) when choosing between Fear Factor foods, $\chi^2(1) = 4.57$, $p < .04$, this preference reversed for desserts, where participants now preferred the high variance option (65%) to the low variance option (35%), $\chi^2(1) = 5.07$, $p < .03$.

Of course, desserts and Fear Factor foods are just two cases of the multitude of positive and negative experiences that people experience. To further buttress our account, we examined preferences for variance for a wider set of experiences by asking an additional group of participants ($N = 56$) about 17 randomly ordered categories (surgeon, birth control, dentist, burglar alarm, car, school, hair dye, washing machine, dorm, vacation, hotel, person to date, coffee, chocolate, CD, book, and movie). Participants reported their preference for outcome variance in each category on a 7-point scale (1: *play it safe* to 7: *take a gamble*), and their perceived valence of each category, also on a 7-point scale (1: *very negative experience* to 7: *very positive experience*). For each category, we then computed the average of both measures for each participant. Confirming the results from Study 2, we observed a highly significant correlation between preference for variance and valence, $r(17) = .60$, $p < .05$. In short, participants preferred lower variance in negatively-valenced domains but higher variance in positively-valenced domains.

STUDY 3: PREFERENCES FOR VARIANCE CONTROLLING FOR RANGE

While the observed preference for high variance in positive domains and low variance in negative domains appears to be robust and in line with our account that extreme reference points lead to convex utility in positive domains and convex utility in negative domains, a related – but distinct – explanation can also account for the results of Study 2. A large body of research has explored attention to outliers in shaping preferences for uncertainty (Budescu et al., 2002; Ganzach, 1995); in this view, the presence of an “outlier” extremely bad or extremely good gamble leads people to overweight such instances, which then shapes preferences for uncertainty. In Study 2, our high variance distributions had greater variance than our low variance distributions, but also a wider range of ratings (see Figure 2.2), allowing for the possibility that outlier ratings present only in high variance options are driving preferences. Our account, however, is that preferences for and aversion to variance are driven by people's different utility functions in positive and negative domains, the result of reference points set by their memories of extreme experiences and not by the presence of extreme ratings in some distribution, and should therefore not be dependent on the presence of outliers. In Study 3, we address this competing explanation by showing that people's preferences for variance hold even when the range of ratings is identical to the range for lower variance options.

Method

Participants (N = 77; 45 female, M_{age} = 23.2) completed this survey online; similar to Study 2, they were asked to choose between two food options based on the ratings these foods had been given by others. The options were described as either desserts or as Fear Factor foods. In this study, we provided a list of 10 ratings for each option on a 10-point scale, one having low variance (SD = 1.84) and one having high variance (SD = 3.31). However, in addition to setting the means of the two options equal at 5.5 as in Study 2, we also set the ranges to be equal by fixing the lowest rating at 2 and the highest rating at 9 for both distributions. Thus the low variance distribution was [2, 4, 5, 5, 5, 6, 6, 6, 7, 9] while the high variance distribution was [2, 2, 2, 3, 3, 8, 8, 9, 9, 9]. Ratings were shown in random order to each participant (see Figure 2.3 for sample).

Figure 2.3

Low Variance	High Variance
5	3
6	2
2	9
7	2
6	8
9	8
4	9
6	9
5	2
5	3

Sample of low and high variance distributions as shown to participants in Study 3. Ratings for the low variance option were randomly drawn from: [2, 4, 5, 5, 5, 6, 6, 6, 7, 9], while ratings for the high variance option were randomly drawn from [2, 2, 2, 3, 3, 8, 8, 9, 9, 9]. Distributions were labeled either "Fear Factor Foods" or "Desserts."

Results and Discussion

We predicted that participants would continue to prefer the high variance dessert and the low variance Fear Factor food, in contrast to the prediction that, if outliers drove preferences in Study 2, participants would be indifferent to variance if the range of ratings was the same.

Results were strikingly similar to those from Study 2, in both direction and magnitude. As before, participants preferred the low variance option (68%) to the high variance option (32%) when choosing between Fear Factor foods, $\chi^2\,(1) = 5.48$, $p < .02$;

this preference reversed for desserts, where they now chose the high variance option (67%) more than the low variance option (33%), χ^2 (1) = 4.00, $p < .05$.

Moderating Variance-Seeking and Variance-Aversion through Attention to Extremes

Having demonstrated utility over experiences consistent with extreme reference points (Study 1) and a resultant tendency toward variance-seeking in positive and variance-aversion in negative domains (Studies 2 and 3), we turn now to examining whether tempering these extreme reference points moderates preferences for variance, as our account suggests it should. As outlined earlier, previous research indicates that promotion-focused individuals process information globally while prevention-focused individuals tend to engage in more local, concrete processing (Forster and Higgins 2005); thus we predicted that a prevention focus might encourage a more balanced weighting of the individual ratings provided by others rather than a reliance on extreme memories as reference points, decreasing participants' sensitivity to variance.

In Studies 4 and 5, we both manipulate regulatory focus and use natural shifts in regulatory focus over life span – using age as a proxy (Heckhausen, 1997; Lockwood et al., 2005) – to demonstrate the impact of decreased reliance on extreme reference points. Because we again use two distributions that differ in variance but have the same mean, if participants engage in more careful processing – in some sense gaining an awareness that the means of the two distributions are identical – then we would expect them to be less sensitive to variance, instead choosing somewhat randomly between two distributions that have the same mean value.

STUDY 4: MANIPULATING REGULATORY FOCUS MODERATES PREFERENCES FOR VARIANCE

Method

A nationally representative sample of participants (N = 456; 219 female, M_{age} = 40.8) completed this survey online as part of a series of unrelated surveys.

We used the same distributions as in Study 2 (see Figure 2.2): Both distributions had the same average rating (M = 5.5) on a 10-point scale, but one was high variance (SD = 2.7) and the other was low variance (SD = 1.1); participants were again randomly assigned to indicate preferences for either desserts or foods from Fear Factor. We counterbalanced whether the high or low variance option appeared first.

In addition, participants were randomly assigned to adopt either a promotion-focused or prevention-focused mindset (Brendl, Higgins, and Lemm, 1995; Cesario, Grant, and Higgins, 2004; Pham and Higgins, 2005). In the promotion condition, they were told that their goal was to "get the best possible dessert" or "get the best possible Fear Factor food." In the prevention condition, in contrast, they were told that their goal was to try to "avoid getting a bad dessert" or "avoid getting a bad Fear Factor food."

Participants indicated which food they would rather eat if asked to choose.

Results and Discussion

Across both conditions, as in previous studies, participants preferred the low variance option (59%) to the high variance option (41%) when choosing between Fear

Factor foods, $\chi^2(1) = 7.37$; this preference reversed for desserts, where they now chose the high variance option (54%) more than the low variance option (46%), $\chi^2(1) = 7.53$, $p < .01$.

Importantly, however, these patterns were impacted by our manipulation of promotion and prevention focus. Participants in a promotion focus showed the same pattern as usual, preferring the low variance option (63%) to the high variance option (37%) when choosing between Fear Factor foods, and a reversal for desserts where they chose the high variance option (57%) more than the low variance option (43%), $\chi^2(1) = 9.22$, $p < .01$. Participants in a prevention focus, who we expected would be less sensitive to variance due to more careful processing, showed no preference for variance, preferring the low variance option just 53% of the time when choosing between Fear Factor foods, and the high variance option just 52% of the time when choosing between desserts, $\chi^2(1) = .54$, $p > .46$.

STUDY 5: AGE MODERATES PREFERENCES FOR VARIANCE

Method

Participants ($N = 179$; 122 female, $M_{age} = 62.6$) completed this survey at the annual AARP meeting in Boston, Massachusetts. We split participants into two groups, following the cut points used by other researchers (Cole et al., forthcoming): “young old” consumers aged 69 and below ($N = 130$; 94 female, $M_{age} = 58.4$) and “old old” consumers aged 70 and above ($N = 46$; 25 female, $M_{age} = 74.7$).

We used the same distributions as in Studies 2 and 4 (see Figure 2.2): Both distributions had the same average rating (*M* = 5.5) on a 10-point scale, but one was high variance (*SD* = 2.7) and the other was low variance (*SD* = 1.1); participants were again randomly assigned to indicate preferences for either desserts or foods from Fear Factor. We counterbalanced whether the high or low variance option appeared first.

Participants indicated which food they would rather eat if asked to choose.

Results and Discussion

Across both age groups, participants once again preferred the low variance option (69%) to the high variance option (31%) when choosing between Fear Factor foods, $\chi^2(1) = 7.37$, $p < .01$, but the high variance option (58%) more than the low variance option (42%) when choosing between desserts, $\chi^2(1) = 3.72$, $p = .054$, though this difference was only marginally significant.

As predicted, however, these preferences were impacted by the age of our participants. For "young old" participants – as with promotion-focused participants in Study 4 – we observed the usual pattern of results, a preference for low variance (76%) over high variance options (24%) when choosing between Fear Factor foods, but a preference for high variance (60%) over low variance options (40%) when choosing between desserts, $\chi^2(1) = 14.02$, $p < .001$. "Old old" participants – as with prevention-focused participants in Study 4 – showed no preference for variance in choice between foods, however; for neither the dessert condition nor the Fear Factor condition were the percentages of "old old" participants choosing the high variance and low variance options significantly different (51% choosing high variance for desserts and 55% choosing high variance for Fear Factor), χ^2 (1) < 1, p = .86.

In sum, while Study 4 offered support for the impact of a prevention focus on decreasing sensitivity to variance via manipulation, Study 5 offers even more evidence for the moderating role of regulatory focus via a "natural experiment," utilizing differences in regulatory focus across the life span. As in Study 4, prevention-focused individuals – in this case, older people – demonstrated decreased sensitivity to variance.

GENERAL DISCUSSION

Taken together, the above studies reveal a great deal about the way that decision-makers process and respond to variance information. We show that, in line with previous research suggesting that reference points for experiences are likely shaped by memories of extreme instances, utility for negative experiences is convex across the range of positive experiences but concave across the range of negative experiences (Study 1). Studies 2 and 3 demonstrate that preferences for high and low variance experiences reflect these utility curves, with people displaying variance-seeking when choosing between positive experiences and variance-aversion when choosing between negative experiences. Study 3 additionally addresses an alternative explanation by validating that the difference in preferences across domains holds even when low and high variance options have the same range of outcomes. Finally, Studies 4 and 5 show that factors which shift focus away from extreme reference points – manipulating regulatory focus and age – can diminish the preferences for variance across domains.

One contribution of this work is to question the intuitively appealing conflation of losses with negative experiences and gains with positive experiences. In fact, we contend that because people set reference points at extreme experiences, most outcomes in negative domains may be treated as gains (leading to risk-averse behavior) while most outcomes in positive domains may be treated as losses (leading to risk-seeking behavior). Of course, not all people set their reference points at the utmost extremes of experience (or we would have observed 100% preference for variance in desserts and 100% aversion to variance in Fear Factor foods): there is likely considerable variability on exactly where people's reference points are defined in both positive and negative domains. Indeed, West and Broniarczyk (1998) demonstrated just such an impact of such more specific reference points in a positive domain in their investigation of consumers reaction to critic reviews of movies. Overall, however, the utility curves that our participants provided in Study 1 are consistent with general variance-seeking in positive domains and variance-aversion in negative domains. Future research is needed to integrate our domain-specific approach (using general preferences across domains with different valences) with an individual-differences approach (taking into account the reference points of specific individuals).

Not All Ratings Are Created Equal

Indeed, the notion of looking more closely at individuals raises interesting questions about how consumers integrate the many different kinds of ratings available on websites into their preferences. We have treated each observation in the distributions that participants observe as equally informative, but of course decision-makers often gather opinions from an unrepresentative sample of the population (Hu, Pavlou, and Zhang, 2005), or place different credibility on different sources of information. First, people may seek out or weight expert ratings more heavily in their judgments than the ratings of random users (Smith, Menon, and Sivakumar, 2005, but see Eliashberg and Shugan, 1997; Fitzsimons and Lehmann, 2004). Second, people frequently seek out or weight more heavily the opinions of people they know (Smith et al., 2005), for example by paying more attention to raters whose preferences seem to match their own (Gershoff, Mukherjee, and Mukhopadhyay, 2007), and ignoring those who are too dissimilar (Yaniv and Milyavsky, 2007). Even in cases when consumers select inferior recommenders by failing to consider their expertise or similarity (Gershoff, Broniarczyk, and West, 2001), collaborative filtering mechanisms – which seek to identify other consumers with similar preferences in order to make recommendations better suited to an individual consumer – may cause people's preferences to be influenced more by similar than dissimilar others (Adomavicius and Tuzhilin, 2005). Thus preferences for variance might be impacted by consumers' views of which *kinds* of consumers – similar or dissimilar others – are creating that variance.

Preferences for Variance in Real Markets

Our experiments involve hypothetical choices between possible options, and we therefore wanted to corroborate the impact of variance using actual market data. We obtained publicly available movie data from Internet Movie Database (IMDb.com), and selected all 4,296 movies with entries for user ratings, opening weekend box office gross, and total box office gross. Ratings were input by IMDb users on a scale from 1 to 10 (whole numbers, 10 being the best) and because the database supplied the total number of ratings for a movie and a 10% range of the percentage of users who gave it each ratings, we were able to extrapolate estimates of the exact number of users who gave each rating, and therefore estimates of the mean and variance of ratings for that movie. In an effort to capture the proportional increase in revenues due to the response of moviegoers to the opinions of other viewers rather than initial marketing efforts, we used the ratio of overall box office gross to opening weekend box office as our dependent measure. Not surprisingly – and an indicator that our metric is valid – movies with higher user ratings performed better at the box office, $\beta = .16$, $p < .001$; more interestingly, variance also served as a positive – and independent – predictor of box office success, $\beta = .07$, $p < .01$.

One possible explanation for these results is that variance is associated with increased word of mouth: for example, movies that are both strongly liked and strongly disliked are more likely to be reviewed, giving them more visibility, leading to increased viewership. While this may account for some of the impact of variance, however, additional analyses suggest it does not account for the entire effect of variance: while total number of ratings was a significant predictor of box office success, $\beta = .05$, $p < .05$, both mean ($\beta = .14$, $p < .001$) and variance ($\beta = .07$, $p < .01$) remained significant and independent predictors. Finally, these data also allow us an additional opportunity to test whether outlier ratings can account for preferences for variance in positive domains, buttressing our results from Study 3, by controlling for extremely positive ratings in this regression. While the number of "10"s a movie received was a positive predictor of box office success ($\beta = .10$, $p < .01$), once again mean ($\beta = .10$, $p < .001$) and most importantly variance ($\beta = .04$, $p < .03$) were significant predictors, suggesting that variance has a positive impact on movie success independent of the presence of outliers.

In sum, although these results cannot be used to tell a causal story of the impact of variance on box office success – unlike the studies we presented earlier – and many factors likely affect the success of high variance movies, they are consistent with our account that the mere divergence of opinions plays a role in determining the appeal of movies in the marketplace.

Given that a product's popularity may be linked to the variance of its ratings on the Internet, how might marketers utilize variance in their marketing efforts? We can imagine scenarios where firms would alter their investments in production and advertising based on the predicted distribution of consumer opinions about an end product. In fact, at least one marketer has already intuited the impact of ratings variance: Jeffrey Kalmikoff, chief creative officer of Threadless.com – which allows users to rate t-shirt designs and produces t-shirts based on those ratings – told the New York Times that in addition to looking at mean ratings, the company also counts the number of 0s and 5s a design receives (on a 0-to-5 scale), stating that designs that "inspire passionate disagreement often get printed because they tend to sell" (Walker, 2007).

When diversity of opinion for other products can be assessed at an early stage – such as pre-screenings of summer comedies, for example – it would behoove movie studios to opt for making potentially higher variance movies instead of lower variance movies. Once a movie was already prepared for release, the distribution of opinions in pre-screening could still help to determine the amount of advertising that would be necessary to achieve a certain target level of popularity; one of the most striking implications of our results is that increasing variance in ratings (by generating both more positive and more negative ratings) might be a better use of marketing dollars than simply trying to increase positive ratings of one's product. Conversely, our findings suggest that in negative domains – for example, preventative health care – where consumers are forced to make "grudge purchases," firms would be wiser to create and advertise products with lower variance in opinion. In short, knowledge of consumer preferences for variance can inform product decisions at every stage from product design to promotional campaigns.

CONCLUSION

Much of the focus of studies of social influence has been on active attempts – by both peers and marketers – to influence opinions and behavior (Cialdini, 2001; Friestad and Wright, 1994), with an emphasis on cases where using the opinions of others leads to errors (Asch, 1951). At the same time, models of preference formation suggest that incorporating the opinions of others leads to more accurate predictions (Clemen and Winkler, 1993). Indeed, the fact that one's preferences have been shaped partly by the preferences of others is one of the reasons it is useful to use one's own preferences to predict the preferences of others (Dawes and Mulford, 1996, Hogarth, 1975). We suggest that understanding the process by which the distribution of opinions across the population impacts individuals involves consideration of not just the average opinion, but also the level of consensus – or, as is often the case, the lack thereof. This aspect of integrating opinions is increasingly important given the proliferation of websites that provide user ratings, rendering variance information both available and in some sense unavoidable. More fundamentally, knowledge about how consumers cope with the uncertainty induced by ratings variance – a more naturalistic and common kind of uncertainty than the monetary gambles frequently studied by social scientists – contributes to a more general understanding of preference formation.

Chapter 3

Whether to Rate and What to Rate? Sample and Response Bias in Online Opinion Sharing

In recent years, online ratings and reviews have become increasingly pervasive within nearly every domain of consumer choice, from selecting books on Amazon.com to selecting hotels on Expedia.com. Whereas in the past, people may have only had the opportunity to ask the opinions of friends and family members (the quality of whose judgments they could assess accordingly), they now have immediate access to the opinions of countless anonymous strangers at their disposal. While there is no question that online opinion forums can serve as a valuable resource – if only for entertainment and community-building – actually making sense of this information presents a new challenge. Users of online ratings must somehow aggregate the abundant and often conflicting opinions of others in order to draw any meaningful conclusions from them.

Indeed, much research has focused on the growing popularity of online ratings and the way that consumers weight these inputs into decision-making. Despite this acknowledgement of the substantial role played by online opinion sharing, though, relatively little attention has been devoted to the study of when and why people rate in the first place. Traditional economic theory predicts that no one would expend the time and effort to rate without some personal gain from doing so, and yet we observe literally billions of ratings on the Internet provided by individuals who get nothing tangible in return. So why do they rate? And more importantly, how does the decision what rating to give depend on the correspondence between the prevailing popular opinion and one's own experience? On the one hand, viewing the ratings of others may impact ex ante product expectations, and thus indirectly change subjective experience and the subsequent rating provided. On the other hand, existing ratings may more directly impact ratings if people wish to either corroborate – or contradict – the opinions of others. If these subtle forces are in effect, the distribution of ratings will fail to represent the frequency of actual experiences across the population, and therefore will serve as a poor guide for users to predict their own outcomes.

In the next section, we survey the literature on online social influence to formulate several hypotheses about the possible determinants of rating bias. We then test our hypotheses in a series of controlled laboratory experiments. Finally, we conclude with a discussion of the information that can be garnered from Internet ratings, and the implications for managers and marketers in using word-of-mouth as tools for communication and product recommendation.

How the Decision to Rate Depends on Experience

We have all found ourselves in situations where we are so delighted or so disgusted with an experience that we simply cannot wait to tell others. In conversation with our immediate acquaintances, we might just be making small talk or hope that our advice will help them and then be reciprocated in the future. However, when someone runs home and logs onto the Internet to blog about their experiences for all the world to see, are similar motivations at play? The short answer is yes. Individuals may feel a desire both to express themselves and to disseminate their personal insights even amongst people they have never met. Although altruism should not be discounted as a key driver of such behavior, raters may also aspire to enhance their social status or gain recognition for their contributions. Hennig-Thurau, Gwinner, Walsh, and Gremler (2004) ascertain four primary incentives for providing electronic word-of-mouth: “desire for social interaction, desire for economic incentives, their concern for other consumers, and the potential to enhance their own self-worth” (p. 39).

Putting aside for the moment economic incentives to fabricate ratings (which we describe further in the Discussion section), this willingness to share private experiences confers a mutual benefit: Specifically, those who share their opinions may gain “feelings of autonomy, competence, relatedness, and value” (Cape, 2007, p. 3). It has also been shown that high-value product knowledge is readily identified by members of online communities, and its transmission helps to build social networks (Dwyer, 2007). There is evidence for a correlation between Internet word-of-mouth and brick-and-mortar sales, suggesting that people either make offline decisions based on online information or that the content of online conversation mirrors that of offline conversation (Godes and Mayzlin, 2004). Of course, the actual value of online product ratings depends on their ability to assist users in making better choices by improving predictions of their own outcomes should they consume the item in question. As such, the usefulness of ratings hinges upon their accurate reflection of the likelihood of realizing various outcomes. Since the decision to rate is conditional on the experience that someone has had, though, people who have had certain sorts of experiences may be more apt to rate. If so, the set of ratings encountered on the average website is a biased, and thus misleading, sample of the range of experiences across the population. (This compounds the bias in those who chose the experience to begin with!)

Consider the example of an individual choosing whether to rate a movie. Unless she is a professional critic or a rating fanatic (a peculiar breed of opinion sharer), she probably does not go to a movie convinced that she will rate it regardless of what she ends up thinking about it. Rather, she exits the theater after the credits roll, her eyes readjust to the daylight, and she formulates her overall impressions. In order to actually supply a rating – and perhaps a detailed verbal review, as well – she must still feel sufficiently passionate about the movie once she returns to her home computer several hours later! This sort of reasoning would generate a greater proclivity to rate amongst those who have had extreme experiences (either positive or negative) than those who have experienced more neutral outcomes. This is precisely what Hu, Pavlou, and Zhang (2006) found in their assessment of reviews on Amazon.com: Scores formed a bimodal distribution of “brag and moan” reviews. Not only does this imply that the middle spectrum of opinions in the population is underrepresented, but further that an asymmetric effect of the two extremes would prevent average scores from converging to true product quality.

The disproportionate tendency to rate extreme experiences may result from their accentuation of the reasons people rate at all. If people rate to express themselves, this may become more salient when staking out an extreme opinion. If people rate for altruism, this may also become more significant when recommending an especially great experience or warning against a horrible experience. Yet another plausible explanation that we have yet to explore is that the ability to impact others increases with extreme ratings. Even before the Internet was a major factor in decision-making, unfavorable ratings were shown to have a greater influence on consumer perceptions of product quality and emotional response than favorable ratings, as a result of attributing negative information to underlying product characteristics and positive information to situational factors (Mizerski, 1982). More recently, it has been shown that the lowest ratings have greater impact than the highest ratings for major online booksellers (Chevalier and Mayzlin, 2006). Price and Stone (2004) even discovered that people prefer advisors who make more extreme judgments because they assume their confidence reflects greater knowledge and accuracy. If this is the case, those who have had extreme experiences might correctly foresee the large impact of their opinions, whereas those with mediocre experiences will not bother to provide an opinion that will be overlooked by users.

How Ratings of Others Shape Expectations and Experience

Having described why people are apparently more inclined to rate extreme experiences, we now take this analysis one step further and examine how the previous ratings of others might influence an individual's rating either by altering her subjective experience or by creating incentives for her to exaggerate. Imagine, for instance, that you selected a restaurant based on the favorable ratings it received on CitySearch.com. Are you more likely to perceive an over-cooked steak and bland vegetable melange as better or worse than if you had never seen the ratings in advance? Alternatively, what if you had seen poor ratings prior to the dinner but had been no choice in eating there for a friend's birthday – then, would you perceive your so-so meal any differently? Research on this matter is unambiguous on one point: Prior expectations can have a huge impact on the way that experiences are perceived. Frequently studied from the perspective of marketers, it is well established that advertising helps to shape consumer hypotheses which they then test through product experience (Hoch and Deighton, 1989). What is less obvious is the direction in which expectations will distort subjective experiences.

Across many situations, there is support for the notion of contrast effects whereby higher expectations lead to lower satisfaction with outcomes, and the reverse. Anderson (1973) demonstrates that product evaluations depend on the disparity between expectations and actual product performance. Likewise, Koszegi and Rabin (2006) develop a model in which the utility of consumption depends not only on outcome-based utility but also on divergence of the outcome from a prior reference point. Differentiating the underlying psychological mechanism according to consumer expertise, Lynch, Chakravarti, and Mitra (1991) show that people with little domain knowledge may actually form mental representations of their experiences relative to prior expectations, whereas individuals with greater knowledge about the ranges of attribute values may be influenced by context only in the way that they anchor response scales. Exploring the topic of ratings in particular, Talwar, Jurca, and Faltings (2007) find that low ratings on TripAdvisor.com are more common following high ratings, and vice versa. They hypothesize that the previous reports influence expectations and a decision-maker’s subsequently divergent experience is assessed in comparative terms.

Despite this compelling evidence for contrast effects, there is the opposite possibility that people who form expectations based on viewing others’ ratings then evaluate experiences more similarly to these expectations than they otherwise would have.

Nickerson (1998) defines this type of confirmation bias as “the seeking or interpreting of evidence in ways that are partial to existing beliefs, expectations, or a hypothesis in hand” (p. 175). Lee, Frederick, and Ariely (2006) conducted an experiment where the timing of information disclosure determined its capacity to alter expectations: Participants who were informed that a beer sample contained several drops on vinegar before tasting it found it less appealing that those who were only told after their sampling. Cialdini and Goldstein (2004) also describe how conformity to the responses of others might be induced by goals such as accurate perception of reality, affiliation with social groups, and positive self-concept. People are more susceptible to interpreting product experience as confirming the messages from advertisers when other evidence remains ambiguous (Hoch and Ha, 1986), so we can extrapolate that they may also be susceptible to the ratings of others when they are less confident in their own evaluations of an experience.

How Both the Decision to Rate and the Rating Given Depend on Ratings of Others

To summarize the research surveyed above, we can draw several conclusions: first, motivations to rate may be enhanced when people have more extreme experiences; and second, exposure to the ratings of others may alter expectations, but it is unclear whether the ensuing subjective experiences will contrast or confirm these expectations, if either. Here, we focus on a related question: How do the decisions whether to rate and what to rate depend on the ratings of others regardless of any influence they may have had prior to experience? In other words, if we presume no effect of ratings on expectations – as in a typical situation where an individual never encountered others’ opinions in advance of experience – can seeing these ratings ex post affect rating?

Consider once again that you had a mildly disappointing meal at a restaurant, and that you visited CitySearch.com afterwards to compare the opinions of others to your own. Supposing that you had never viewed the ratings there before your dining experience, it seems unlikely that they would alter your opinion at this point – given that you would have already articulated it so well to your dinner companions – but these ratings might nonetheless affect your conviction and thus your desire to contribute your own two cents!

The result, we contend, is that a larger discrepancy between your opinion and existing ratings will induce you to share your own rating, and furthermore, induce you to overcompensate for the "error" in others' judgments by exaggerating your own rating. The basis for this argument is the "false consensus effect" originally put forth by Ross, Greene, and House (1977). The authors demonstrate the robust phenomenon by which individuals perceive their own opinions as more common and representative of the population at large than the opinions of others. This behavior persists even in circumstances where greater utilization of others' advice objectively improves decision-making (Yaniv, 2004; Yaniv and Milyavsky, 2007). In the domain at hand, such overconfidence may accentuate the motivation to rate when one's opinion diverges from the sentiments expressed by others since the latter are perceived to be less predictive for future visitors to the website. In an act of self-expression or altruism, or some combination thereof, an individual would certainly wish to assert her opinion, and perhaps even more vehemently than she would have in the absence of the "inaccurate" ratings in order to bring the average rating (prominently displayed on many websites) closer to her own view. Returning to the restaurant example, we hypothesize that observing rave reviews from others on CitySearch.com would increase your adamancy about the shortcomings of the restaurant and inspire you not only to rate, but perhaps to rate more negatively than was even warranted by your subjective experience!

STUDY OF RATING PROVISION

Applying the aforementioned literature to the domain of online ratings, we propose three main hypotheses: (1) People are more inclined to rate extreme experiences than moderate experiences; (2) People are more inclined to rate experiences that do not correspond to the current average opinion; and (3) People whose opinions do not correspond to the current average tend to over-compensate by providing ratings even further from the average in the same direction as their "true" opinion. Note that, unlike the explanation for sequential rating fluctuations provided by Talwar, Jurca, and Faltings (2007), these hypotheses do not rest on any assumption about the impact of ex ante exposure to ratings on subjective experience because the theoretical evidence is ambiguous regarding contrast versus conformity effects.

We test our hypotheses in a single laboratory experiment where we show participants pairs of photographs of other people and ask them both to choose which they would like to rate, and subsequently to provide their rating for the attractiveness of the chosen photo. Across conditions, we varied whether participants saw average ratings (sometimes actual and sometimes contrived by us) given to the photos by others to see how these influenced their own decisions whether and what to rate.

Pre-Test

We obtained a database of photographs from the website HotOrNot.com which allows visitors to rate photos of people on a scale from 1: not to 10: hot. With several hundred photos, we merely used the average ratings from the website to select the 10 highest and 10 lowest rated photos for each females and males, for a total of 40 photos. Because we could not be sure how closely ratings on the HotOrNot website would correspond to the ratings of our subject pool, we asked 45 participants in an unrelated lab study to rate these photos on a similar 10-point scale ranging from 1: extremely unattractive to 10: extremely attractive.

Based on the average ratings of our participants, we divided the photos of females and males separately into four groups labeled “L” (low), “ML” (medium-low), “MH” (medium-high), and “H” (high), with the rating cutoffs defined so that each group contained exactly five photos. The average rating ranges for each group (which were different for female and male photos) are shown in Table 3.1.

Table 3.1: Average rating ranges for the five photos in each category

	Female	Male
L	2.3 - 3.0	2.3 - 3.4
ML	3.0 - 4.8	3.5 - 4.0
MH	5.6 - 6.6	4.9 - 5.7
H	6.8 - 7.2	5.9 - 6.8

For each gender, there was a significant pair-wise difference in ratings between photos in the “L” and “MH” categories, and between photos in the “ML” and “H” categories, so we could safely conclude that all five photos in the latter categories are considered relatively more attractive than all five photos in the respective former categories.

Method

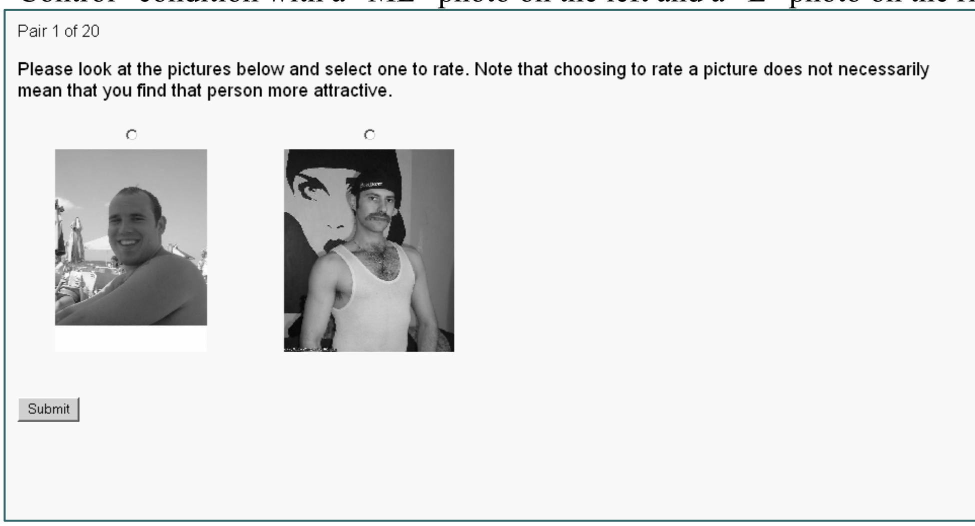
Having established these photo categories in the pre-test, the main goal of the study was to see which photos people would be most inclined to rate when forced to choose, and how this choice and subsequent rating would be affected by seeing the ratings that others had supposedly already provided.

32 participants (15 female; age $M = 21.9$, $SD = 4.23$) took part in this study. They were each shown 10 pairs of female photos and 10 pairs of male photos in random order, and with random pairings so that each pair contained either one photo from the "L" category and one from the "ML" category, or alternatively one photo from the "MH" category and one from the "H" category. In other words, every participant saw all of the 40 photos used in the pre-test in random same-sex pairings of "L" with "ML" and "MH" with "H". Participants were also randomized into one of two conditions, either the "Control" condition where they simply chose one of the two photos to rate and rated it or the "Ratings Shown" condition where they saw average ratings supposedly given by other people to each photo before choosing and rating. In all cases, the ratings shown for the "L" and "H" photos were the actual pre-test averages, but for each "ML" photo a random rating in the high range was shown and for each "MH" photo a random rating in the low range was shown in order to assess how choices and ratings were affected by a perceived error in judgment by previous raters. Because the differences in average pre-test ratings were significantly between both the "ML" and "H" and the "L" and "MH"

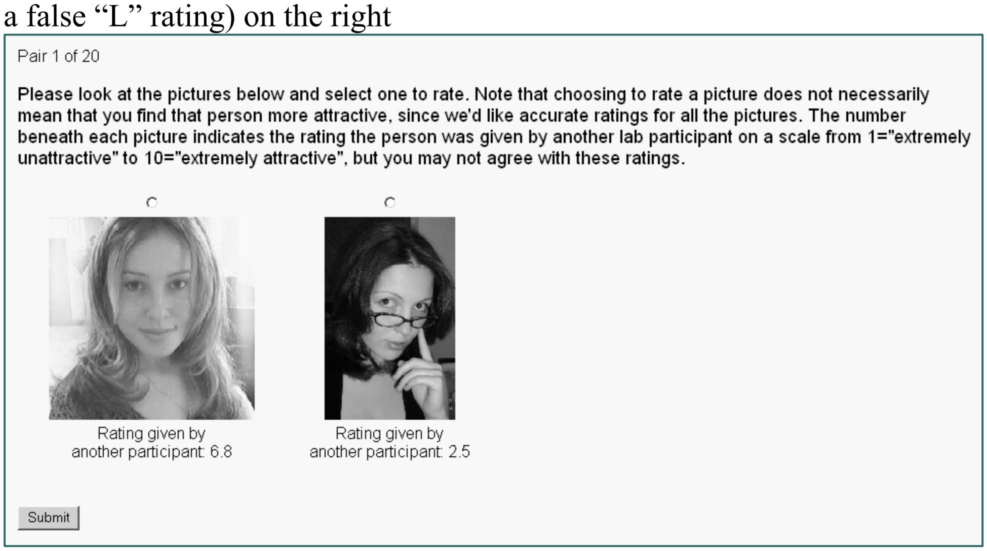
categories, we can be confident that these false ratings provided would be viewed as inaccurate by our participants here. (See Figure 3.1 for sample screenshots seen by participants in each condition.)

Figure 3.1

“Control” condition with a “ML” photo on the left and a “L” photo on the right



“Ratings Shown” condition with a “H” photo on the left and a “MH” photo (with a false “L” rating) on the right



Sample screenshots seen by participants in making the choice which photo to rate.

To test the first hypothesis that people are inclined to rate more extreme experiences, we looked just at the "Control" condition to see whether more than 50% – expected by pure chance – of the choices participants made were to rate the "L" photo (instead of the paired "ML" photo) or to rate the "H" photo (instead of the paired "MH" photo). To test the second hypothesis that people are inclined to rate experiences that seem to be misrepresented by the prevailing opinion, we looked at just the "Ratings Shown" condition to assess – oppositely – whether more than half of participants chose to rate the moderate photos of each pair, as these were the photos for which the average ratings shown were inaccurate. Finally, to test the third hypothesis that people would over-compensate for such perceived errors, we isolated out analysis to just the "Ratings Shown" condition choices where participants did, in fact, choose the moderate option, and assessed whether they tended to decrease their rating even further below the pre-test ratings for "ML" photos that had purportedly been given high average ratings from others and ever further above the pre-test ratings for "MH" photos that had purportedly received a low average rating from others.

Results

Our hypotheses were partially supported by the experimental data. As predicted, participants in the "Control" condition were more inclined to rate the extreme of the two photo options (60.5% of choices) than were participants in the "Ratings Shown" condition (52.7% of choices), as the latter were driven toward the more moderate options for which false ratings were shown, $t(638) = 1.97$, $p < 0.05$. However, while these choices represent more frequent extreme choices for the “Control” participants than would be expected from random chance, $X^2(1) = 16.4$, $p < 0.001$, obviously choices for the “Ratings Shown” participants are not fewer than half. We suspect that the desire of participants to rate options for which their opinions differed from the current opinion merely dampened the competing tendency to rate extremes, without actually reversing it.

The results also failed to support our third hypothesis that those who did choose to rate the moderate options in the “Ratings Shown” condition would overcompensate for the supposed error of others by providing lower ratings for “ML” photos (with false “H” ratings shown) and higher ratings for “MH” photos (with false “L” ratings shown) as compared to those in the “Control” condition. Instead, we found that ratings of those in the control actually conformed somewhat to the false ratings shown: The average rating given to “ML” photos was 4.40 in the “Ratings Shown” versus 4.21 in the “Control” condition, though not a significant difference, $t(171) = 0.749$, $p = 0.455$. The difference in average rating given to “MH” photos was significant: 5.11 in the “Ratings Shown” versus 6.08 in the “Control” condition, $t(98) = 2.51$, $p = 0.0136$. We infer that the desire to conform to social standards dominates a desire for self-expression in this domain. Certainly, one of the main proposed drivers for overcompensation was the desire to correct the average as part of an altruistic incentive to help others make better choices, but this incentive was sorely lacking in the chosen domain.

DISCUSSION

One of the most fascinating aspects of any online network is the interdependency in behaviors of participants. In the case of opinion sharing that we have examined here, the ratings of others shape expectations, consumption choices, and even subsequent rating decisions. In accordance with prior research (Hu, Pavlou, and Zhang 2006), we find that individuals are more apt to share opinions when they feel more strongly, and further, that they are more apt to share opinions that diverge from the current consensus. In the above study, these competing motives tempered one another, but there are certainly instances where the opposite might be true such that more extreme opinions could also be ones that contradicted the consensus, and we predict that motives to rate would be reinforced in these circumstances.

Unlike Talwar, Jurca, and Faltings (2007), we do not propose modified expectations and a subsequent contrast effect in experience as reason for the tendency to share divergent opinions, but rather a realization of the greater impact that one's opinion will have in situations where it differs from the consensus. To test this more precisely, future studies will compare rating provision across conditions in which participants see ratings of others before or after their own experiences. We predict that expectations will be impacted more when individuals see ratings in advance of forming their own opinions (and even then their experiences are liable to confirm rather than contrast expectations), but argue that this is not the case for many online raters. This assessment is supported by the research of Herr, Kardes, and Kim (1991) showing that word-of-mouth has a greater impact when opinions are less strongly formulated, and Fitzsimons and Lehmann (2004) showing that individuals with established opinions accept only concordant recommendations of experts even though they end up with lower satisfaction as a result.

Nonetheless, if the expectations of decision-makers are not altered by the ratings of others in advance of experience, their own ratings may still be influenced by seeing others' ratings *after* their experience. In addition to studying decisions about which experiences to rate, we hypothesized that the altruistic and self-expression motives to rate would induce overcompensation in divergent ratings to correct the perceived error of prior ratings. Although we found the opposite that participants in our study modified their opinions toward agreement with the existing ratings, the suggested psychological mechanism may merely have been absent from the domain of photo attractiveness. In future studies, it will be important to choose a domain where, first, participants believe that their ratings are going to assist others in making better consumption choices, and second, participants understand the influence that their rating will have on the average rating seen by future consumers. This is more realistic to the majority of online rating websites, and may enhance motives to bring average rating as close to possible as one's own experience.

Distribution Bias and Diagnosticity

Describing individual biases in rating provision begs the question of how the overall distribution of ratings is possibly distorted. If raters indeed see ratings of others in advance of experience and discover confirmatory evidence of their expectations, then the earliest ratings might have a disproportionate impact such that subsequent ratings cluster around them more so than independent opinions across the population otherwise would.

Alternatively, it is plausible that raters who form strong opinions of their own based on experience overcompensate in their ratings – as we predict – leading to more dispersed ratings than would reflect differences in subjective experience alone. It is beyond the scope of this research to analyze the many potential outcomes of this dynamic evolution of ratings distributions, yet we raise it as an interesting topic for future research, and one that may be informed by computer simulations to calibrate the effect of changing parameters in rating behavior on broader community-wide trends.

Another related topic of interest is whether users of online ratings become aware of these distributional distortions and re-weight the value they place on both individual ratings and the distribution as a whole. While it would be a tough proposition for decision-makers to backwards induce the diagnosticity of each rating based on the biases described above – and there may be additional concerns of deliberate deception of raters seeking to promote their products (Lam and Riedl, 2004) – users of ratings may at least understand that some of these forces are in play and thus treat all ratings as noisy signals of another person's possibly unrepresentative viewpoint (Mayzlin, 2006). Even so, this is a setting in which both the probabilities of outcomes and the outcomes themselves remain ambiguous to decision-makers, given the diversity of subjective experience across raters (Einhorn and Hogarth, 1986; Hogarth, 1975). Hsee and Hastie (2006) summarize several general decision-making biases that may lead people to choose options that fail to maximize their ultimate satisfaction, and we suspect that each of these may be exacerbated for offline choices made via online information such as ratings. For instance, the impact bias by which people overestimate the impact that an affective experience will have could lend extreme ratings undue influence. The projection bias and distinction bias are especially relevant when people make decisions in a different state of arousal or in a different evaluation mode, respectively, than they are at the time of consumption. These biases may affect the quality of decision-making as well as post-experience evaluations and ratings.

Enhancing the Usefulness of Ratings

Despite widespread biases in rating provision indicated by both prior research and the study results reported here, ratings of others still have great potential to assist in improving one's decisions. One promising avenue for enhancing the usefulness of opinion sharing is measurement of the diagnosticity of specific raters. It has been shown that the average online shopper searches only a very restricted number of e-commerce websites (Johnson, Moe, Fader, Bellman, and Lohse, 2004), and may be equally circumscribed in considering a subset of the innumerable ratings available. Therefore, in order to realize greater value from ratings, it is helpful to make more prominent those ratings that will best predict someone's own experience within a particular product category (Gershoff, Broniarczyk, and West, 2001). If decision-makers have access to multiple ratings of a given rater, they may take into account both the presumed knowledge of the rater and the correlation of the rater's prior ratings with their own past experiences (Budescu, Rantilla, Yu, and Karelitz, 2003).

Personalized recommendations can further reduce the burden of search on decision-makers (Smith, Menon, and Sivakumar, 2005), especially when automated through collaborative and content –based filtering that take into account, respectively, the similarity in tastes between individuals and similarity in options matching a given individual's preferences (Ansari, Essegaier, and Kohli, 2000). Tracking the reputation of raters may also provide clues as to their general expertise and credibility (Leskovec, Adamic, and Huberman, 2006; (White, 2005), and the accuracy of raters may improve via monetary rewards (Sniezek, Schrah, and Dalal, 2004). Although Achrol and Kotler describe such developments as a shift in marketing "from being the agent of the seller to being an agent of the buyer" (1999, p. 146), new technologies that aid in decision-making have also been shown to increase sales at e-commerce sites (Schafer, Konstan, and Riedl, 1999). Despite a required tradeoff between the simplicity of information and the value of its content, both desirable properties of web search and choice can be retained through transparent filtering and sorting of information, with the end goal of encouraging a more optimal weighting by users.

CONCLUSIONS

The objective of this research is to develop a more comprehensive framework for describing the way that opinions are exchanged online. One piece of the puzzle is to understand when and why individuals choose to rate in the first place, a behavior that contradicts purely rational economic theory. As we have elaborated, the conditioning of incentives to rate on both the extremity of experience and the divergence of experience from others may lead to distortions in the set of opinions that are provided. Furthermore, the rating provided may not be independent of others' prior ratings either if these ratings have altered expectations in advance of subjective experience (as proposed by other researchers), or if their presence alters the value that a given rating will have for the community (as we highlight here). How these dynamics will play out in shaping the distribution of ratings is unclear, though future work will strive to illuminate the most important features of such technology-mediated online interactions.

FOLLOW-UP STUDY A

This study will test similar hypotheses to those of the study reported above, but with a slightly modified design to overcome some of the previous limitations. First, we use the domain of humor articles rather than photo attractiveness to overcome to some extent the tendency for people to choose whichever option they prefer as opposed to the one that they would be more inclined to rate in the real world. While people share opinions for a number of reasons in daily life, the paradigm of our study to use choice between options to extrapolate comparative likelihood of rating is problematic if participants feel that their choice reflects preferences and wish to express themselves in that way. We suspect that participants may be especially prone to this misinterpretation in the rating of photo attractiveness since they are likely quite familiar with selecting other individuals that they find more attractive. In addition, there is little motivation to assist others in making good decisions through accurate photo ratings. For humor articles, however, it will hopefully seem more plausible that others will use the ratings in making a choice for which options to read, as well as providing recognition to the authors with ratings that reflect true quality. Here, participants will also be told that the purpose of providing ratings is to rank order the articles according to the average humor ratings that they have received. We hope that this will inspire participants to strive for accuracy in average ratings not just amongst the very best articles (which are those that others would encounter first in the real world) but also amongst the worst articles.

Method

As for the last study, we will run a pre-test on a set of humor articles to assess baseline evaluations of subject pool participants on a 10-point scale 1: *not at all funny* to 10: *extremely funny*. Although these ratings will not be compared directly to ratings of participants in the main study in case of systematic differences between participants in the different sessions, they will be used to sort the articles into categories with low ("L"), medium-low ("ML"), medium-high ("MH") and high ("H") ratings with the cutoffs set so that equal numbers fall within each category.

Participants in the main study will again see a series of random pairings of humor articles – "L" with "MH" and "ML" with "H" – and be asked to choose which to rate and then provide a rating on the same 10-point scale. As for the preceding study, they either see no ratings of others associated with the articles (the "Control" condition) or see ratings supposedly given by others. Now, however, the latter will be further randomized into one of two conditions "Ratings Before" or "Ratings After", in which participants will be exposed to the ratings of others either before or after they read the articles themselves. We excluded a simultaneous exposure condition (as for the photos) since we expected that the ready visibility of a number would before the reading of an article would be equivalent to the "Ratings Before" condition. Also, we wanted to distinguish between the *ex ante* impact of others' ratings on expectations and subjective experience versus the *ex post* impact of others' ratings on motivations to rate in a certain way. As for the HotOrNot study with photos, the "L" and "H" articles will be shown with their actual pre-test average ratings whereas the "ML" articles will be shown with false ratings drawn randomly from the high range and "MH" articles will be shown with false ratings drawn randomly from the low range.

We hypothesize, first, that people will be more inclined to rate the "L" (instead of "MH") and "H" (instead of "ML") articles in the "Control" condition since participants will feel more strongly about their quality of humor. Second, we hypothesize that more participants will choose to rate the moderate "ML" and "MH" articles (with false ratings shown) in both the "Ratings Before" and "Ratings After" conditions than in the "Control" condition, though we do not necessarily predict a complete reversal in favor of rating these since the proclivity to rate extremes will remain a competing motivation. Finally, we hypothesize that amongst those who chose to rate the more moderate articles, ratings given will conform to the false ratings of others in the "Ratings Before" condition, but contrast the false ratings of others in the "Ratings After" condition. Assuming that the ratings of those in the "Control" are not significantly different from pre-test ratings, the ordering of ratings would then be "Ratings Before" < "Control" < "Ratings After" for the "MH" articles with false low ratings, and "Ratings After" < "Control" < "Ratings Before" for the "ML" articles with false high ratings.

FOLLOW-UP STUDY B

This study is designed to better understand the psychological mechanisms that generate the observed tendency to share opinions that are extreme relative to the opinions of others, as well as the tendency to overcompensate in ratings to bring the average closer to subjective reality. Several forces at work – possibly in tandem – are self-expression and altruism. Although these may be indistinguishable in practice if personal opinions claimed online are always observable to others (and it is believed that others are behooved to adopt one's own views), we can disentangle them to some extent in the lab by assessing the separate effects on rating behavior due to enhancing one or both motivations in isolation.

Method

Using the same stimuli of humor articles as in Study A, we now manipulate the self-expression and altruistic motivations to rate. The paradigm is similar to the "Ratings After" condition of Study A in that articles are paired "L" with "MH" and "ML" with "H", and the ratings of others shown after the articles are pre-test actual for both "L" and "H", but random false low ratings for "MH" and random false high ratings for "ML". Holding constant the timing of exposure to others' ratings, we can now assess the way that these ratings of others interact with incentives to rate. Here, the "Low Motivation" condition will be identical to the "Ratings After" condition in Study A, with a stated goal of deriving rankings for the humor articles based on their average ratings. The "Self-Expression" condition will request a brief explanation for each rating and emphasize that one's sense of humor will be judged based on his shared opinions. Finally, the "Altruism" condition will inform participants that articles will be priced to future readers based on their average rating (from 1 to 10) in cents.

We predict that both self-expression and altruism are intrinsic incentives to rate, so that enhancing the salience of either will accentuate the effects that we observe in the "Ratings After" condition of Study A. Thus, compared to the "Low Motivation" condition, we expect a greater selection of moderate articles for which false ratings are shown ("MH" instead of "L" and "ML" instead of "H") in both the "Self-Expression" and "Altruism" conditions. Furthermore, we expect that amongst those choices of moderate articles, ratings will over-compensate more in both the "Self-Expression" and "Altruism" conditions, such that ratings for "ML" articles (with false high ratings shown) will be lower and ratings for "MH" articles (with false low ratings shown) will be higher than for the "Low Motivation" condition. We make no prediction about whether these effects will be stronger for "Self-Expression" versus "Altruism" since the degree to which we are able to enhance these factors is highly specific to the manipulations, and there may in fact be some overlap in the two.

Summary of Contributions

Drawing upon the diverse disciplines of Behavioral Economics, Computer Science, Marketing, and Psychology, the research projects described above explore several ways that people make sense of abundant or conflicting information, particularly when decisions are mediated by Internet technologies. The chapters focused, respectively, on the way that information is disproportionately weighted based on its presentation, content salience, and relevance to others. One emergent theme is that choices in such information-rich environments are highly context-dependent and subject to systematic biases, perhaps subconsciously by those purveying or using the information. This topic warrants further attention so that the valuable information on the web can be put to better use through more appropriate filtering, organization, and display.

The specific objective of the first paper was to demonstrate that preferences are not fixed, but rather learned, constructed, and modified according to the organization of information. In replication of prior research, we show that individuals explicitly bestow greater importance to option attributes that are given a greater "share" of the categorization, but taking this one step further, we also show that such weightings are implicitly reflected in choices between options. Unlike past research that has examined "partition dependence" in the spreading of probability estimates across groupings of events or consumption across groupings of options, we instead emphasize the distribution of weights across attributes. Both decision-makers and information purveyors may falsely assume that the importance of such factors in predetermined, but in fact, preferences can be redirected simply through the categorization of attributes. The significance of this experimental result is evident in the seemingly arbitrary way that real websites partition option attributes.

Given the vast quantity of information online, it is frequently helpful for users to rely on summary statistics, but the tradeoff for simplicity is obviously a loss of some informational content. In particular, the second paper looked at whether the full distribution of online ratings is meaningful to individuals above and beyond the average rating. We corroborate existing research in the finding that variance affects preferences, but differently depending on circumstance. Whereas other research has attended to the significance of exceeding some aspiration level, we suggest that positive experiences should not be falsely conflated with “gains” and negative experiences with “losses” relative to a null reference point. The salience of extreme outcomes is heightened by association with more readily recalled experiences, leading to greater expected likelihood and hedonic impact of realized these extremes. We propose this as a mechanism to explain the observed pattern of preference for high variance options in positive domains where decision-makers strive to achieve the best possible outcomes and for low variance options in negative domains where decision-makers aim to avoid the worst possible outcomes. On the whole, this research delves into the ways that variance in subjective experiences across individuals differs from uncertainty in monetary gambles.

The third paper takes a novel approach to exploring when and what opinions people provide, thus leading to the potentially unrepresentative distribution of ratings observed on a typical website. While past research sheds light on various stages of online decision-making, there has been little work elaborating on the holistic process by which individual expectations, subjective experiences, and motivations to rate are, in turn, influenced by the opinions of others. Here, we explore both sample bias in the decision to provide ratings and response bias in the subsequent decision of which rating to provide. While we confirm prior research indicating that people are more likely to share opinions about experiences that are more extreme in some absolute sense, we also propose that self-expression and altruistic incentives to rate may be accentuated when one's opinion differs more from the prevailing average opinion. In addition, we challenge the assumption that exposure to others' opinions must alter expectations in advance of experience in order to impact ratings, since others' opinions could also impact the value of supplying a particular rating *after* the experience has already taken place. We consider these various effects in conjunction to better understand the dynamic evolution of real ratings distributions on the Internet.

Each of these projects attempts to unify themes of information aggregation that span multiple disciplines of behavioral research. Although core theories of individual decision-making isolate the effects of particular contextual and social variables, they make no strong predictions about preference formation and choice in more complicated online networks where many factors simultaneously affect the exchange of information. The ongoing research discussed here aims to disambiguate the most salient influences on decision-makers when they are confronted with large amounts of varied information, and ultimately to assist people in correcting systematic biases that generate suboptimal decisions. These topics will become increasingly important as interaction across the Internet expands to new aspects of daily life in the coming years. Not only will people consciously seek out information online, but it will also be critical for web-based applications to intelligently distill meaning on behalf of users.

References

Achrol, Ravi S., and Philip Kotler (1999). "Marketing in the Network Economy." *Journal of Marketing*, 63, 146-163.

Adomavicius, Gediminas, and Alexander Tuzhilin (2005). "Personalization Technologies: A Process-Oriented Perspective," *Communications of the ACM*, 48 (10), 83-90.

Anderson, Rolph E. (1973). "Consumer Dissatisfaction: The Effect of Disconfirmed Expectancy on Perceived Product Performance." *Journal of Marketing Research*, 10 (1), 38-44.

Ansari, Asim, Skander Essegaier, and Rajeev Kohli (2000). "Internet Recommendation Systems." *Journal of Marketing Research*, 37 (3), 363-375.

Ariely, Dan (2008). Predictably Irrational. New York, N.Y.: HarperCollins Publishers.

Ariely, D., and M. I. Norton (2008). "How Actions Create – Not Just Reveal – Preferences." *Trends in Cognitive Sciences*, 12, 13-16.

Asch, Solomon E. (1951), "Effects of Group Pressure upon the Modification and Distortion of Judgments," in H. Guetzkow (ed.), *Groups, Leadership, and Men*, Pittsburgh, PA: Carnegie Press, 177-190.

Bakos, Yannis J. (1997). "Reducing Buyer Search Costs: Implications for Electronic Marketplaces." *Management Science*, 43 (12), 1676-1692.

Baumeister, Roy F., Ellen Bratslavsky, E., Catrin Finkenauer, and Kathleen D. Vohs (2001). "Bad is Stronger than Good." *Review of General Psychology*, 5, 323–370.

Benartzi, S., and R. H. Thaler (2001). "Naive Diversification Strategies in Retirement Saving Plans." *American Economic Review*, 91 (1), 79–98.

Benbasat, I., and P. Todd (1992). "The Use of Information in Decision Making: An Experimental Investigation of the Impact of Computer-Based Decision Aids." *MIS Quarterly*, 16 (3), 373-393.

Bettman, James R. (1973). "Perceived Risk and Its Components." *Journal of Marketing Research*, 10 (2), 184-190.

Bettman, James R., and Pradeep Kakkar (1977). "Effects of Information Presentation Format on Consumer Information Acquisition Strategies." *Journal of Consumer Research*, 3 (4), 233-240.

Bettman, J. R., M. F. Luce, and J. W. Payne (1998). “Constructive Consumer Choice Processes.” *Journal of Consumer Research*, 25, 187-217.

Brendl, C. Miguel, E. Tory Higgins, and Kristi M. Lemm (1995). “Sensitivity to Varying Gains and Losses: The Role of Self-Discrepancies and Event Framing.” *Journal of Personality and Social Psychology*, 69 (6), 1028-1051.

Budescu, David V., Kristine M. Kuhn, Karen M. Kramer, and Timothy R. Johnson (2002). “Modeling Certainty Equivalents for Imprecise Gambles.” *Organizational Behavior and Human Decision Processes*, 88 (2), 748-768.

Budescu, D.V., A.K. Rantilla, H. Yu, and T.M. Karelitz (2003). “The Effects of Asymmetry among Advisors on the Aggregation of their Opinions.” *Organizational Behavior and Human Decision Processes*, 90, 178-194.

Cape, Peter (2007). “Understanding Respondent Motivation.” *Survey Sampling International*, 2007.

Cesario, Joseph, Heidi Grant, and Tory Higgins (2004). “Regulatory Fit and Persuasion: Transfer from ‘Feeling Right’.” *Journal of Personality and Social Psychology*, 86 (3), 388-404.

Chakravarti, A., and C. Janiszewski (2003). “The Influence of Macro-Level Motives on Consideration Set Composition in Novel Purchase Situations.” *Journal of Consumer Research*, 30, 244-258.

Chakravarti, Amitav, Chris Janiszewski, and Gulden Ulkumen (2006). “The Neglect of Prescreening Information.” *Journal of Marketing Research*, 43, 642-653.

Chevalier, Judith A. and Dina Mayzlin (2006). “The Effect of Word of Mouth on Sales: Online Book Reviews.” *Journal of Marketing Research*, 43, 345-354.

Cialdini, Robert B. (2001), *Influence: Science and Practice*, Needham Heights, MA: Allyn and Bacon.

Cialdini, Robert B., and Noah J. Goldstein (2004). “Social Influence: Compliance and Conformity.” *Annual Review of Psychology*, 55, 591-621.

Clemen, Robert T., and Robert L. Winkler (1993). “Aggregating Point Estimates: A Flexible Modeling Approach.” *Management Science*, 39 (4), 501-515.

Dawes, R. M., and M. Mulford (1996). “The False Consensus Effect and Overconfidence: Flaws in Judgment or Flaws in How We Study Judgment?” *Organizational Behavior and Human Decision Processes*, 65 (3), 201-211.

Diehl, Kristin (2005). "When Two Rights Make a Wrong: Searching Too Much in Ordered Environments." *Journal of Marketing Research*, 42, 313-322.

Diehl, K., L. J. Kornish, and J. G. Lynch (2003). "Smart Agents: When Lower Search Costs for Quality Information Increase Price Sensitivity." *Journal of Consumer Research*, 30, 56-71.

Diehl, Kristin, and Gal Zauberman (2005). "Searching Ordered Sets: Evaluations from Sequences under Search." *Journal of Consumer Research*, 31, 824-832.

Dwyer, Paul (2007). "Measuring the Value of Electronic Word of Mouth and Its Impact in Consumer Communities." *Journal of Interactive Marketing*, 21 (2), 63-79.

Einhorn, Hillel J., and Robin M. Hogarth (1986). "Decision Making Under Ambiguity." *The Journal of Business*, 59 (4), Part 2: The Behavioral Foundations of Economic Theory, pp. S225-S250.

Eliashberg, Jehoshua, and Steven M. Shugan (1997). "Film Critics: Influencers or Predictors?" *Journal of Marketing*, 61(2), 68-78.

Fitzsimons, Gavan J., and Donald R. Lehmann (2004). "Reactance to Recommendations: When Unsolicited Advice Yields Contrary Responses." *Marketing Science*, 23 (1), 82-94.

Forster, Jens, and E. Tory Higgins (2005). "How Global Versus Local Perception Fits Regulatory Focus." *Psychological Science*, 16 (8), 631-636.

Fox, Craig R., David Bardolet, and Daniel Lieb (2005). "Partition Dependence in Decision Analysis, Resource Allocation, and Consumer Choice" in R. Zwick and A. Rappoport (eds.), Experimental Business Research, Vol. 3. Dordrecht, The Netherlands: Kluwer Academic Publishers, 338-360.

Fox, Craig R., and Robert T. Clemen (2005). "Subjective Probability Assessment in Decision Analysis: Partition Dependence and Bias toward the Ignorance Prior." *Management Science*, 51 (9), 1417-1432.

Fox, Craig R., Rebecca K. Ratner, and Daniel S. Lieb (2005). "How Subjective Grouping of Options Influences Choice and Allocation: Diversification Bias and the Phenomenon of Partition Dependence." *Journal of Experimental Psychology*, 134 (4), 538-551.

Fox, Craig R., and Yuval Rottenstreich (2003). "Partition Priming in Judgment Under Uncertainty." *Psychological Science*, 14 (3), 195-200.

Friestad, Marian, and Peter Wright (1994). "The Persuasion Knowledge Model: How People Cope with Persuasion Attempts. *Journal of Consumer Research*, 21, 1-31.

Frost, J. H., Z. Chance, M. I. Norton, and D. Ariely (in press). "People are Experience Goods: Improving Online Dating with Virtual Dates." *Journal of Interactive Marketing*.

Ganzach, Yoav (1995). "Attribute Scatter and Decision Outcome: Judgment versus Choice." *Organizational Behavior and Human Decision Processes*, 62 (1), 113-122.

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), 418-438.

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., Ashesh Mukherjee, and Anirban Mukhopadhyay (2003). "Consumer Acceptance of Online Agent Advice: Extremity and Positivity Effects." *Journal of Consumer Psychology*, 13 (1&2), 161-170.

Godes, David, and Dina Mayzlin (2004). "Using Online Conversations to Study Word-of-Mouth Communication." *Marketing Science*, 23 (4), 545-560.

Hearst, Marti A. (2006). "Clustering Versus Faceted Categories for Information Exploration." *Communications of the ACM*, 49 (4), 59-61.

Heckhausen, Jutta (1997). "Developmental Regulation across Adulthood: Primary and Secondary Control of Age-Related Challenges." *Developmental Psychology*, 33 (1), 176-187.

Hennig-Thurau, Thorsten, Kevin P. Gwinner, Gianfranco Walsh, and Dwayne D. Gremler (2004). "Electronic Word-of-Mouth via Consumer-Opinion Platforms: What Motivates Consumers to Articulate Themselves on the Internet?" *Journal of Interactive Marketing,* 18 (1), 38-54.

Herr, Paul M., Frank R. Kardes, and John Kim (1991). "Effects of Word-of-Mouth and Product-Attribute Information on Persuasion: An Accessibility-Diagnosticity Perspective." *Journal of Consumer Research*, 17 (4), 454-462.

Higgins, E. Tory (1997). "Beyond Pleasure and Pain." *American Psychologist*, 52 (12), 1280-1300.

Hirsh, Haym, Chumki Basu, and Brian D. Davison (2000). "Learning to Personalize." *Communications of the* ACM, 43 (8), 102-106.

Hoch, Stephen J., and John Deighton (1989). “Managing What Consumers Learn from Experience.” *Journal of Marketing*, 53 (2), 1-20.

Hoch, Stephen J., and Young-Won Ha (1986). “Consumer Learning: Advertising and the Ambiguity of Product Experience.” *Journal of Consumer Research*, 13 (2), 221-233.

Hogarth, Robin M. (1975). “Cognitive Processes and the Assessment of Subjective Probability Distributions.” *Journal of the American Statistical Association*, 70 (350), 271-289.

Hsee, Christopher K., and Reid Hastie (2006). “Decision and Experience: Why don’t We Choose What Makes Us Happy?” *TRENDS in Cognitive Science*, 110 (1).

Hu, Nan, Paul A. Pavlou, and Jennifer Zhang (2006). “Can Online Reviews Reveal a Product’s True Quality? Empirical Findings and Analytical Modeling of Online Word-of-Mouth Communication.” *Proceedings of the 7th ACM Conference on Electronic Commerce*, Ann Arbor, MI, 324-330.

Huber, J., and N. Kline (1991). “Adapting Cutoffs to the Choice Environment: The Effects of Attribute Correlation and Reliability.” *Journal of Consumer Research*, 18, 346-357.

Johnson, Eric J., Wendy W. Moe, Peter S. Fader, Steven Bellman, and Gerald L. Lohse (2004). “On the Depth and Dynamics of Online Search Behavior.” *Management Science*, 50 (3), 299-308.

Kahneman, Daniel, and Amos Tversky (1979). “Prospect Theory: An Analysis of Decision under Risk.” *Econometrica*, 47 (2), 263-292.

Kleinmuntz, D., and D. Schkade (1993). “Information Displays and Decision Processes. *Psychological Science*, 4, 221-227.

Koszegi, Botond, and Matthew Rabin (2006). “A Model of Reference Dependent Preferences.” *Quarterly Journal of Economics*, 121 (4), 1133-1165.

Lam, Shyong K., and John Riedl (2004). “Shilling Recommender Systems for Fun and Profit.” *Proceedings of the 13th International Conference on World Wide Web*, New York, NY, 393-402.

Lee, Leonard, Shane Frederick, and Dan Ariely (2006). “Try It, You'll Like It: The Influence of Expectation, Consumption, and Revelation on Preferences for Beer.” *Psychological Science*, 17 (12, 1054-1058.

Leskovec, Jurij, Lada A. Adamic, and Bernardo A. Huberman (2006). “The Dynamics of Viral Marketing” *Proceedings of the 7th ACM Conference on Electronic Commerce*, Ann Arbor, MI, 228-237.

Lockwood, Penelope, Alison L. Chasteen, and Carol Wang (2005). "Age and Regulatory Focus Determine Preferences for Health-Related Role Models." *Psychology and Aging*, 20 (3), 376-389.

Lurie, N. H., and C. H. Mason (2007). "Visual Representation: Implications for Decision Making. *Journal of Marketing*, 71, 160-177.

Lynch, John G., Jr., Dipankar Chakravarti, and Anusree Mitra (1991). "Contrast Effects in Consumer Judgments: Changes in Mental Representations or in the Anchoring of Rating Scales?" *Journal of Consumer Research*, 18 (3), 284-297.

Mandel, Naomi, and Eric J. Johnson (2002). "When Web Pages Influence Choice: Effects of Visual Primes on Experts and Novices." *Journal of Consumer Research*, 29, 235-245.

March, James G., and Zur Shapira (1992). "Variable Risk Preferences and the Focus of Attention." *Psychological Review*, 99 (1), 172-183.

Marsden, Peter V., and Noah E. Friedkin (1993). "Network Studies of Social Influence." *Sociological Methods and Research*, 22 (1), 127-151.

Mayzlin, Dina (2006). "Promotional Chat on the Internet." *Marketing Science*, 25 (2), 155-163.

Mizerski, Richard W. (1982). "An Attribution Explanation of the Disproportionate Influence of Unfavorable Information." *Journal of Consumer Research*, 9 (3), 301-310.

Mogilner, C., T. Rudnick, and S. Iyengar (forthcoming). "The Mere Categorization Effect: How the Presence of Categories Increases Choosers' Perceptions of Assortment Variety and Outcome Satisfaction." *Journal of Consumer Research*.

Morewedge, Carey K., Daniel T. Gilbert, and Timothy D. Wilson (2005). "The Least Likely of Times." *Psychological Science*, 16 (8), 626-630.

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

Norton, M. I., J. H. Frost, and D. Ariely (2007). "Less is More: The Lure of Ambiguity, or Why Familiarity Breeds Contempt." *Journal of Personality and Social Psychology*, 92, 97-105.

Novemsky, Nathan, and Ravi Dhar (2005). "Goal Fulfillment and Goal Targets in Sequential Choice." *Journal of Consumer Research*, 32 (3), 396-404.

Ogilvie, Daniel M., Kristin M. Rose, and Jessica B. Heppen (2001). "A Comparison of Personal Project Motives in Three Age Groups." *Basic and Applied Social Psychology*, 23 (3), 207-215.

Payne, J. W., J. R. Bettman, and E. J. Johnson (1992). "Behavioral Decision Research: A Constructive Processing Perspective." *Annual Review of Psychology*, 43, 87-131.

Payne, John W., Dan J. Laughhunn, and Roy Crum (1980). "Translation of Gambles and Aspiration Level Effects in Risky Choice Behavior." *Management Science*, 26 (10), 1039-1060.

Peeters, Guido and Janusz Czapinski (1990). "Positive–negative Asymmetry in Evaluations: The Distinction between Affective and Informational Negativity Effects." In W. Stroebe & M. Hewstone (Eds.). *European review of social psychology* (Vol. 1, pp. 33–60). New York: Wiley.

Pham, Michel Tuan, and E. Tory Higgins (2005). "Promotion and Prevention in Consumer Decision-Making" in S. Ratneshwar and David Glen Mick (eds.), *Inside Consumption: Consumer Motives, Goals, and Desires.* Routledge, UK, pp. 8-43.

Prelec, Drazen, Birger Wernerfelt, and Florian Zettelmeyer (1997). "The Role of Inference in Context Effects: Inferring What You Want from What Is Available." *Journal of Consumer Research*, 24 (1), 118-125.

Price, Paul C., and Eric R. Stone (2004). "Intuitive Evaluation of Likelihood Judgment Producers: Evidence for a Confidence Heuristic." *Journal of Behavioral Decision Making*, 17 (1), 39-57.

Pu, Pearl, and Li Chen (2006). "Trust Building with Explanation Interfaces." *Proceedings of the 11th International Conference on Intelligent User Interfaces*, Sydney, Australia, 93-100.

Ratner, Rebecca K., Barbara E. Kahn, and Daniel Kahneman (1999). "Choosing Less-Preferred Experiences for the Sake of Variety." *Journal of Consumer Research*, 26 (1), 1-15.

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

Read, Daniel, George Loewenstein, and Matthew Rabin (1999). "Choice Bracketing." *Journal of Risk and Uncertainty*, 19 (1-3), 171-197.

Roberts, J. H., and J. M. Lattin (1991). "Development and Testing of a Model of Consideration Set Composition." *Journal of Marketing Research*, 28, 429-440.

Ross, Lee, David Greene, and Pamela House (1977). "The 'False Consensus Effect': An Egocentric Bias in Social Perception and Attribution Processes." *Journal of Experimental Social Psychology*, 13 (3), 279-301.

Salganik, Matthew J., Peter Sheridan Dodds, and Duncan J. Watts (2006). "Experimental Study of Inequality and Unpredictability in an Artificial Cultural Market." *Science*, 311 (5762), 854-856.

Schafer, J. Ben, Joseph Konstan, and John Riedl (1999). "Recommender Systems in E-Commerce." *Proceedings of the 1st ACM Conference on Electronic Commerce,* Denver, CO, 158-166.

Simonson, Itamar (1990). "The Effect of Purchase Quantity and Timing on Variety-Seeking Behavior." *Journal of Marketing Research*, 27 (2), 150-162.

Simonson, Itamar, and Russell S. Winer (1992). "The Influence of Purchase Quantity and Display Format on Consumer Preference for Variety." *Journal of Consumer Research*, 19 (1), 133-138.

Smith, Donnavieve, Satya Menon, and K. Sivakumar (2005). "Online Peer and Editorial Recommendations, Trust, and Choice in Virtual Markets." *Journal of Interactive Marketing,* 19 (3), 15-37.

Sniezek, Janet A., Gunnar E. Schrah, and Reeshad S. Dalal (2004). "Improving Judgment with Prepaid Expert Advice." *Journal of Behavioral Decision Making*, 17 (3, 173-90.

Spiekerman, Sarah and Corina Paraschiv (2002). "Motivating Human-Agent Interaction: Transfering Insights from Behavioral Marketing to Interface Design." *Electronic Commerce Research*, 2 (3), 255-285.

Stewart, Neil, Nick Chater, and Gordon D.A. Brown (2006). "Decision by Sampling." *Cognitive Psychology,* 53 (1), 1-26.

Talwar, Arjun, Radu Jurca, and Boi Faltings (2007). "Understanding User Behavior in Online Feedback Reporting." *Proceedings of the 8rd ACM Conference on Electronic Commerce*, San Diego, CA, 134-142.

Thaler, Richard H., and Eric J. Johnson (1990). "Gambling with the House Money and Trying to Break Even: The Effects of Prior Outcomes on Risky Choice." *Management Science*, 36 (6), 643-660.

Todd, P., and I. Benbasat (1994). "The Influence of Decision Aids on Choice Strategies: An Experimental Analysis of the Role of Cognitive Effort." *Organizational Behavior and Human Decision Processes*, 60 (1), 36-74.

Tversky, Amos (1972). "Elimination by Aspects: A Theory of Choice." *Psychological Review*, 79 (4), 281-299.

Walker, Rob (2007, July 8). "Mass Appeal." *The New York Times Magazine*.

West, Patricia M. (1996). "Predicting Preferences: An Examination of Agent Learning." *Journal of Consumer Research*, 23 (1), 68-80.

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.

White, Tiffany Barnett (2005). "Consumer Trust and Advice Acceptance: The Moderating Roles of Benevolence, Expertise, and Negative Emotions." *Journal of Consumer Psychology*, 15 (2), 141-148.

Yaniv, Ilan (2004). "Receiving Other People's Advice: Influence and Benefit." *Organizational Behavior and Human Decision Processes*, 93, 1-13.

Yaniv, Ilan, and Maxim Milyavsky (2007). "Using Advice from Multiple Sources to Revise and Improve Judgments." *Organizational Behavior and Human Decision Processes*, 103 (1), 104-120.

Zhang, S., and G. J. Fitzsimons (1999). "Choice-Process Satisfaction: The Influence of Attribute Alignability and Choice Limitation." *Organizational Behavior and Human Decision Processes*, 77, 192-214.