

# UC Riverside

## UC Riverside Electronic Theses and Dissertations

### Title

Accidentally Bayesian: Preference Similarity Effects on Advice Taking

### Permalink

<https://escholarship.org/uc/item/4zh3x56c>

### Author

Shen, Hang

### Publication Date

2016

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA  
RIVERSIDE

Accidentally Bayesian: Preference Similarity Effects on Advice Taking

A Dissertation submitted in partial satisfaction  
of the requirements for the degree of

Doctor of Philosophy

in

Management

by

Hang Shen

December 2016

Dissertation Committee:

Dr. Ye Li, Chairperson

Dr. Amnon Rapoport

Dr. Rami Zwick

Copyright by  
Hang Shen  
2016

The Dissertation of Hang Shen is approved:

---

---

---

Committee Chairperson

University of California, Riverside

## Acknowledgements

This work would not have been possible without the support of my parents, Dr. Jinwen Shen and Dr. Xiaojie Zhu. I am delighted to share with them the joy of accomplishment.

Nobody has been more important to me in the pursuit of this project than my Fiancée, Min Yu, whose love and support are always with me.

My special thanks to Dr. Rami Zwick and Dr. Amnon Rapoport, for their continuous help for five years. They brought me into this wonderful world of academia in the first place.

I am especially indebted to Dr. Ye Li, Chairperson of my Dissertation Committee, my advisor, and a great human being, who has been supportive of my career goals and who worked actively to guide me academically and personally. He has shown me, by his example, what a good scientist (and person) should be.

## ABSTRACT OF THE DISSERTATION

Accidentally Bayesian: Preference Similarity Effects on Advice Taking

by

Hang Shen

Doctor of Philosophy, Graduate Program in Management

University of California, Riverside, December 2016

Dr. Ye Li, Chairperson

Consumers increasingly depend on online word-of-mouth to inform their purchase decisions. Thanks to the recent trend toward adding social network features to online recommendation platforms and vice versa, potential consumers can now see reviewers' profiles and rating histories and use that information to seek out advice from users with similar preferences. We propose that consumers taking advice from a reviewer will do so according to how similar the reviewer's preferences seem to their own, based on the degree to which they both liked or both disliked the same products or experiences in the past. We show that people make two systematic errors when making affective forecasts about potential products in the presence of preference similarity information. First, they tend to underestimate the degree of preference similarity with the reviewer, especially for mutual disliking. Second, they tend to overweigh her advice relative to a Bayesian advice-taker, particularly for negative advice. These errors bias affective forecasts in opposite directions, even cancelling each other out for negative advice, and can lead to surprisingly accurate forecasts.

## Table of Contents

Introduction .....	1
Hypotheses .....	11
Overview of Studies .....	17
Yelp Field Study .....	17
Study 1 .....	21
Study 2 .....	26
Study 3 .....	34
Study 4 .....	40
General Discussion .....	48
Figures and Tables .....	53
References .....	60

## Lists of Figures

Figure 1. Experimental paradigm used across all studies.....	53
Figure 2A. Distributions of number of common reviews between non-friends .....	54
Figure 2B. Distributions of number of common reviews between friends.....	54
Figure 3A. Distributions of fuzzy matching $PM_1$ among non-friends.....	55
Figure 3B. Distributions of fuzzy matching $PM_1$ among friends.....	55
Figure 3C. Distributions of fuzzy matching $PM_2$ among non-friends.....	56
Figure 3D. Distributions of fuzzy matching $PM_2$ among friends.....	56



## **Lists of Tables**

Table 1. Results for Study 1.....	57
Table 2. Results for Study 2.....	57
Table 3. Results for Study 3.....	58
Table 4. Results for Study 4.....	59

## *INTRODUCTON*

Consumers today have access to far more information when making purchase decisions than ever before, due to the proliferation of consumer-oriented websites and apps. When consumers consider buying a new product, service, or experience, they may try to predict how much they will like it—i.e., make an affective forecast. To better inform their forecast, consumers may read product descriptions and images or rely on word-of-mouth (WOM) advice by reading product reviews on websites such as Yelp, Amazon, and IMDb.

A growing body of WOM research has studied online reviews and their impact on sales (Chen, Wang, and Xie 2011; Chen and Xie 2008; Chevalier and Mayzlin 2006; Duan, Gu, and Whinston 2008; Godes and Mayzlin 2009; Liu 2006; Zhu and Zhang 2010). Research has shown that online WOM is an effective marketing channel because consumers in online communities are particularly likely to share their views of products and services (Brown, Broderick, and Lee 2007). However, research has yet to explore the impact of reviews becoming more social, a trend that has accelerated in recent years. E.g., Yelp allows its users to add other users as friends on its own community or through a Facebook connection. On the other hand, social media websites such as Facebook and Twitter now provide users with product suggestions based on their friends' ratings and activity. This socialized WOM marketing is a more effective communication tool than traditional marketing strategies (Trusov, Bucklin, and Pauwels 2009).

From the consumer's perspective, one effect of online reviews becoming more social is that people reading reviews can now learn more about the reviewers they are receiving

WOM advice from, including information about the reviewer's tastes and preferences. However, research has yet to examine whether this preference information affects how much consumers rely on advice and whether it helps them make more accurate affective forecasts (He and Bond 2013). The present research examines how consumers make affective forecasts using WOM advice when given information about the adviser's (i.e., the reviewer's) preferences. That is, does knowledge about an adviser's preferences affect how consumers use their advice? If a consumer knows an adviser's preference are 90% (vs. 60%) matched with her own (e.g., a close friend vs. a stranger), how will this information determine her acceptance of the adviser's WOM advice and her affective forecasts based on the advice?

To answer this question, we draw from research on advice taking, affective forecasting, and Bayesian updating. In one field study and four experiments, we investigate how people incorporate advice from an adviser with high or low preference similarity into their affective forecasts. To assess the *normative* degree to which consumers should take advice from an adviser in the presence of PM information, we introduce a Bayesian framework that allows us to evaluate forecast accuracy in a quantified way. We find that compared to a Bayesian consumer, people display *false uniqueness*—they exaggerate how different their preferences are from the adviser's—and simultaneously rely too much on WOM advice. Because these errors have opposing effects, people's affective forecasts were surprisingly close to Bayesian, especially for negative advice. We thus extend the literature on advice taking and WOM to better understand advice taking in the social media era.

### *Improving Consumer Affective Forecasting Using WOM Advice*

When choosing between products and experiences, consumers are often guided by how they anticipate the consumption experience will make them feel. Accurate *affective forecasting* therefore seems prerequisite to making good purchase decisions; yet, a large body of literature has found that people's ability to predict their future feelings is far from perfect (e.g., Loewenstein, O'Donoghue, and Rabin 2003; Wilson and Gilbert 2005; Patrick, MacInnis, and Whan Park 2007). For example, people systematically overestimate the impact of future events (Gilbert et al. 1998; Kermer et al. 2006; Mellers and McGraw 2001) and consequently fail to make decisions that maximize their future happiness (Gilbert and Ebert 2002). Systematic errors can also occur when people are unable to apply others' experiences to their own (Walsh and Ayton 2009) or focus too much on current and distinctive features (focusing illusion; Schkade and Kahneman 1998; Wilson et al. 2000).

Although people may often be poor affective forecasters, recent research has suggested that people can more accurately forecast their liking of an experience (e.g., a speed date) by basing their forecast on how a "surrogate" (e.g., someone who had gone on a speed date with the same person) rated the experience (Gilbert et al. 2009)—i.e., WOM. Relying on a surrogate's advice led people to make significantly better affective forecasts than people who made forecasts based on a description of the experience (e.g., their dating profile). Yet, when given the choice between reading a written description and receiving a surrogate's advice, the majority of people preferred the former. The

preference for reading descriptions over receiving advice contrasts with the fact that many consumers *do* actually seek and use advice from online WOM to inform purchase decisions (Gershoff, Mukherjee, and Mukhopadhyay 2003; Mudambi and Schuff 2010). We address this apparent gap in the literature in this article.

Similarly, research on advice-taking has robustly demonstrated that people do not rely much on advice when making quantitative judgments (Bonaccio and Dalal 2006; Sniezek, Schrah, and Dalal 2004; Yaniv, Choshen-Hillel, and Milyavsky 2011; Yaniv and Milyavsky 2007; Yaniv 2004). That is, people put too much weight on their own opinion or private information and insufficient weight on the information others provide. Although egocentrism in advice-taking is well established, this literature has focused exclusively on judgments of objective quantities (e.g., estimating the length of the Nile River or the probability that a randomly drawn ball will be red) and largely ignored the arguably more pervasive process of advice taking for subjective experiences (e.g., food and entertainment). Again, this aversion to advice is inconsistent with the fact that consumers often do seek out and are influenced by WOM advice. It appears that advice taking for matters of taste may differ from advice taking for objective, quantitative judgments.

### *Preference Similarity*

In particular, one potential explanation for the apparent contradiction between people's aversion to advice in the lab and consumer behavior in the real world is that people may have beliefs about the usefulness of the advice. For example, an intuitive way to quantify the usefulness of advice in matters of taste is to consider the preference

similarity between the consumer and the adviser. Taking advice from anonymous strangers online *should* differ from taking advice from someone familiar due to knowing more information about the latter's preferences. Consumers may find it more useful to take advice from someone with similar preferences than from someone with dissimilar preferences, at least for matters of taste.

When the advice comes from an adviser with known preferences, consumers must consider both the valence of the advice and how informative that advice is based on the preference similarity between the consumer and the adviser. For example, perhaps the consumer and adviser often watch movies together and therefore know each other's movie preferences. Alternatively, websites now make it possible for consumers to click on other users' profiles and see what other ratings they have provided. Specifically, consumers can consider their *preference matching* (PM) with the adviser in that domain. PM is our simple operationalization of preference similarity as the percentage of products in a domain (e.g., movies) that the consumer and adviser have similar preferences for (i.e., mutually liked or mutually disliked). We can further distinguish between PM+ and PM- as PM for mutual liking and mutual disliking, respectively. Below, we will derive the exact degree to which consumers should take advice based on their PM with the adviser.

Despite its relevance for advice taking, preference similarity between the advisee (consumer) and adviser (reviewer) has received little attention in the advice taking literature due to its tendency to focus on objective judgments, for which preferences are irrelevant. Although some research has explored how adviser-advisee similarity impacts

usage of WOM (e.g., Price, Feick, and Higie 1989; Naylor, Lamberton, and Norton 2011; He and Bond 2015; Gino, Shang, and Croson 2009), the focus has been on recommender characteristics (gender, expertise, etc.), which are sometimes referred to as “membership group similarity” or “profile similarity” (He and Bond 2013; Simons, Berkowitz, and Moyer 1970; Yaniv, Choshen-Hillel, and Milyavsky 2011). Yaniv and colleagues (2011) distinguished between “profile similarity” and “behavioral similarity” as important components of similarity comparisons. These two categories of similarity differ in their effects on personal attitudes and judgments (Simons, Berkowitz, and Moyer 1970) and may also be distinct in their impacts on advice-taking.

The impact of preference similarity with a specific adviser on WOM usage has been relatively unexplored (Eggleston et al. 2015). This lack of attention is surprising, considering how important preference similarity seems for WOM recommendations (Bonhard and Sasse 2006), especially when the profile characteristics of reviewers are often not easily identifiable or relatively poor predictors of preference similarity. For example, not all middle-aged white males like the same sports teams, music, or restaurants.

Consumers are likely to have some intuitive idea about how similar their friends’ preferences are based on shared experiences or products they both own, but many websites have made it increasingly possible to see this information even for strangers online. For example, preference similarity cues are saliently displayed on Douban (a Chinese movie, book, and music rating website with over 200 million users) in the form of a list of mutually liked items. IMDb, the leading movie rating website, directly

provides users with a preference similarity statistic showing how much their movie ratings overlap with another user's. Along with this similarity statistic, IMDb also lists similarly and dissimilarly rated movies, as well as movies the user may like based on preference similarity. The social music recommendation website Last.fm creates a public profile for each user showing their tastes based on listening history as well as a “taste-o-meter” with a scale from “very low” to “super” that indexes the “musical compatibility” between any pair of users. More generally, *collaborative filtering* is a common technique used by nearly all recommender systems (e.g., Amazon and Netflix) to recommend products based on a user's past preferences based on other users with similar preferences (Linden, Smith, and York 2003; Su and Khoshgoftaar 2009).

Given the proliferation of preference information on WOM websites, this article formally examines how well consumers incorporate implicit or explicit preference similarity information into their judgments to make more accurate forecasts about future consumption experiences. Our studies therefore manipulated the PM between participants and an adviser and investigated how this information impacts people's affective forecasts about how likely they are to like a future consumption experience—i.e., probabilistic affective forecasting (Yates et al. 1996).

#### *Advice Taking Paradigm*

To present our hypotheses with sufficient context, we first introduce the paradigm (See Figure 1) our experiments used to mimic how a consumer might use advice to make a probabilistic affective forecast about a potential purchase (e.g., how likely will I like or dislike this movie). First, participants self-reported their general preference toward



products in a domain and then rated a series of baseline products in that domain. We told them about another consumer who had rated the same products (the adviser). We manipulated PM as the percentage of baseline products the participant's and adviser's ratings agreed on (i.e., both "liked" or "somewhat liked", or both "disliked" or "somewhat disliked"). Finally, we showed participants the adviser's rating of the target product (the advice) and asked for their probabilistic affective forecast for liking the target product.

### *Bayesian Framework*

We evaluate *ex-ante* forecasting accuracy using Bayesian updating as the normative. According to Bayes' Rule, consumers must consider two variables when incorporating advice into their forecasted enjoyment of a future experience: 1) their general liking of products in this domain (i.e., their prior or *base rate*), and 2) the informational value of the advice, which is equivalent to PM in this setting. By combining these two pieces of information into an updated belief using Bayes' Rule, we can calculate a Bayesian forecast of the likelihood that the consumer will like or dislike the future experience and compare this to the consumer's actual (probabilistic) affective forecast.

Importantly, our focus on *ex-ante* forecasting errors contrasts with previous research focusing on *ex-post* errors—comparing the consumer's forecast with her actual experience (e.g., Wilson and Gilbert 2003; Kermer et al. 2006; He and Bond 2013). We focus on *ex-ante* forecasting errors because they are less affected by individual variation in taste and cannot be affected by potential carryover effects of the forecast on the actual consumption experience.

In addition, whereas most prior research on consumer purchase decisions focused on external factors such as price, friends, and advertisements, relatively few studies have discussed the influence of consumers' past experience with the product or service. Our adoption of the Bayesian framework allows us to capture both external (adviser's PM and advice) and internal factors (prior experience with the product domain). Moreover, although the Bayesian beliefs in previous studies usually referred to an objective fact (e.g., whether a target taxi is blue or green), less is known about beliefs that are formed from our own experiences and how similar the advisor's taste compares to ours (Scott and Yalch 1980).

To make this more concrete, consider a consumer, Lily, deciding whether to try a new Chinese restaurant based on a recommendation from a coworker who tried the restaurant last week. Lily has a base rate for liking Chinese restaurants based on past experiences and she needs to consider the usefulness of her coworker's experience (PM could be 70%) or a stranger's Yelp review (PM could be 50%). Optimally combining one's base rate with a friend's or online reviewer's recommendation requires Bayesian updating. Ultimately, advice taking is not just deciding whether to take the advice or not, but about how to combine the advice with prior experience to make an accurate affective forecast.

Formally, let  $P(L)$  denote the consumer's base rate of liking products in a domain. Then,  $P(L|A)$ , the probability of Lily liking the product ( $L$ ) given that the adviser liked it ( $A$ ), is given by Bayes' rule:

$$P(L|A) = \frac{P(A|L) \times P(L)}{P(A)}$$

, where  $P(A|L)$  is the probability of the adviser liking products that Lily liked and  $P(A) = P(L) \times P(A|L) + P(L') \times P(A|L')$ , where  $L'$  and  $A'$  denote Lily and the adviser not liking the product, respectively. PM occurs when both people like or dislike the product, so we formally define preference matching as:<sup>1</sup>

$$PM = P(L \cap A) + P(L' \cap A') = P(A|L) \times P(L) + P(A'|L') \times P(L')$$

For example, suppose that Lily reports that she likes 80% of Chinese restaurants in general. With no further information, she should forecast an 80% likelihood of liking any particular Chinese restaurant. Now, suppose that she receives the positive WOM advice about the restaurant and has 70% PM with the adviser: her ratings for 10 other Chinese restaurants agree with the adviser's ratings for 7 restaurants and disagree for the rest. If Lily is Bayesian, she should be 90% confident about liking the new restaurant:

$$P(L|A) = \frac{70\% * 80\%}{80\% * 70\% + (1 - 80\%) * (1 - 70\%)} = 90.3\%$$

However, suppose that Lily instead perceives the PM with her coworker to be 55%. Assuming that Lily is still a Bayesian forecaster despite underestimating PM, her forecast would now update much less:

$$P(L|A) = \frac{55\% * 80\%}{80\% * 55\% + (1 - 80\%) * (1 - 55\%)} = 83\%$$

Using the Bayesian affective forecast as a normative standard of comparison, we are interested in how people combine their prior preferences with high or low PM advice to form a forecast for liking a target product compared with the Bayesian forecast.

---

<sup>1</sup> Note that while  $PM+ = P(A|L)$  and  $PM- = P(A|L')$ , the relationship between PM with PM+ and PM- is less straightforward. However, in practice, we expect that people will treat PM similarly as PM+ and PM-.

## *HYPOTHESES*

### *Part 1: Perceived Preference Similarity*

A first step to accurate advice taking is that people accurately perceive PM information (PPM). Although we expect subjective PPM to relate to objective PM, there is reason to believe that the relationship between PPM and PM will be imperfect, with previous research having identified errors in both directions.

Specifically, work on the false consensus effect suggests that people think others have similar preferences, whereas work on false uniqueness suggests the opposite. On the one hand, the false consensus effect suggests that people overestimate self-other similarity for a variety of opinions and abilities (e.g., Marks and Miller 1987; Ross, Greene, and House 1977). On the other hand, a false uniqueness effect has been found for desirable abilities people consider themselves strong on and for personally important opinions (Campbell 1986; Suls and Wan 1987). These effects make opposing predictions in terms of how people perceive PM information and it is unclear whether users of review websites will focus on a reviewer's similar ratings as an indication of shared commonality or on a reviewer's divergent ratings as a sign of different preferences.

In the consumer WOM context, Naylor and colleagues has shown that people generally infer that ambiguous reviewers have preferences similar to their own, at least compared to a dissimilar reviewer (Naylor et al. 2011). However, their studies found that participants rated both reviewers with similar and ambiguous profiles at 5 out of 9 in terms of preference similarity. That is, even similar reviewers were seen as only moderately similar, suggesting that for matters of taste, a false uniqueness effect may be

more likely in the consumer WOM context. Furthermore, their studies manipulated profile similarity (gender, age, and geographic location), whereas we provide participants with direct preference similarity information, which may further amplify the false uniqueness of WOM advisers. (In the General Discussion, we further discuss the relationship between profile similarity and preference similarity.)

A number of mechanisms further support the prevalence of false uniqueness in the consumer WOM context. First, many consumers generally have a need to perceive themselves as unique. Research on consumer need-for-uniqueness has found that consumers have a desire to be different from others (Tian, Bearden, and Hunter 2001), because perceiving the self to be undistinguished from others can be seen as a threat to self-identity and have negative emotional impacts (Lynn and Harris 1997; Snyder and Fromkin 1977). Second, negativity bias in perception and memory (Kanouse 1984; Skowronski and Carlston 1987) may cause people to overweigh instances of disagreement over instances of agreement when evaluating PM, especially if judgments are from memory. Third, people naturally focus on their own preferences as a reference point for judging self-other similarity, and this egocentrism can cause people to judge self-other differences as larger than they actually are (Mussweiler 2001; Srull and Gaelick 1983). This focus on the self is a natural focal point in consumer advice taking because ultimately, the consumer is making a forecast about their own liking.

Taken together, false uniqueness, consumers' desire for uniqueness, negativity bias, and egocentrism all contribute to consumers overestimating the unique of their preferences and thus underestimating preference similarity with advisers compared to the

objective level of PM. These mechanisms also suggest that this bias toward perceived dissimilarity will increase with actual PM. That is, people's perceptions of PM will be biased toward a low prior.

**H1:** Consumers assume dissimilarity of preferences with others as their prior, such that they perceive PM (PPM) to be lower than actual PM when PM is high than this prior ( $PPM < PM$  when PM is high).

In addition people underestimating PM, we also expect an effect of valence based on research showing that people believe that they are particularly different from others in terms of what they dislike (Gershoff, Mukherjee, and Mukhopadhyay 2008, 2007). Gershoff and colleagues explained that hated attributes are perceived to be more ambiguous than loved ones. That is, consumers may like products for similar reasons but dislike them for different reasons. In perceiving self-other similarity (i.e., PM), we therefore expect that people perceive mutual disliking to be less common than mutual liking.

**H2:** Consumers perceive PM to be lower for mutual dislikes than mutual likes ( $PPM- < PPM+$ ).

### *Part 2: Advice Taking and Preference Similarity*

Using advice accurately also requires that people can use PM information to determine how to appropriately weigh the advice relative to their own baseline preferences. Past research on advice-taking has found many factors that influence advice utilization for matters of fact (for review, see Bonaccio and Dalal 2006) including similarity between the advisee and adviser (He and Bond 2013; Meng, Chen, and Bartels

2016; Price, Feick, and Higie 1989; Suls, Martin, and Wheeler 2002, 2000; Twyman, Harvey, and Harries 2008; Yaniv, Choshen-Hillel, and Milyavsky 2011). Specifically, past opinion agreement (i.e. behavioral similarity) between the consumer and the advisor has been found to impact advice-taking decisions (Gershoff, Mukherjee, and Mukhopadhyay 2003; Suls, Martin, and Wheeler 2002). Similarity with others also predicts whether people view others as trustworthy (Brewer 1979; DeBruine 2002; Mahajan and Wynn 2012), which then contributes to a higher acceptance of their advice (Gino and Schweitzer 2008; Sniezek and Van Swol 2001).

We therefore predict that consumers are more willing to take advice from advisers whose preferences are similar to their own.

**H3:** Consumers' affective forecasts will be more influenced by WOM advice from an adviser when the adviser has higher PM with them.

#### *Combining Preference Base-Rates, Advice, and Preference Similarity*

Next, we compare actual advice taking to the Bayesian standard. Even if PPM is accurately perceived, people can still fail to appropriately apply Bayes' rule. Earlier work on whether people are Bayesian focused on judgments about objective quantities and used externally provided statistics (Ajzen and Fishbein 1975; Scott and Yalch 1980; Trope and Burnstein 1975; Trope 1974), as opposed to judgments about matters of taste with endogenous base rates. The most common finding in this literature is *base-rate neglect*—people tend to overweigh new information and underweigh or even completely neglect the base rate of occurrence. Given the widespread documentation of base-rate neglect (Bar-Hillel 1980; Kahneman and Tversky 1973; Tversky and Kahneman 1974), it

is possible that people also make affective forecasts that overweigh the adviser's WOM advice while underweighing their own base rate preferences.

However, there is also reason to believe that base-rate neglect may not generalize to advice taking on matters of taste. Whereas most studies that find base-rate neglect provided participants with exogenous base rates, advice-taking decisions are distinguished by the fact that they require people to estimate their own endogenous base rate by recalling past experiences. This difference may be critical: Studies that let participants learn base rates from direct experience rather than explicitly stated probabilities often do not find base-rate neglect (Goodie and Fantino 1999; Medin and Edelson 1988), perhaps because base rates are more salient when learned from experience. In the advice-taking context, the consumer's personal preferences are likely to be particularly salient. The WOM setting is further differentiated by the fact that the diagnosticity of the advice (i.e., PM with the adviser) may also be implicitly learned from direct experience.

Although there is evidence pointing both directions, the main driver for base-rate neglect is that the new information (i.e., the advice) is more salient and that fact remains true when taking WOM advice, especially in cases where there is little other relevant product information available (often true in matters of taste). We therefore expect that people will generally overweigh advice in their probabilistic affective forecasts:

**H4:** People overweigh WOM advice (i.e., adviser's rating) and underweigh their own base rate preferences (relative to Bayesian updating) when making affective forecasts.



Note that hypotheses H1 and H4 make opposite predictions in terms of how much people will be influenced by WOM advice. They exaggerate the uniqueness of their own preferences but nonetheless overweigh the advice from the adviser with supposedly different preferences. Since these errors bias affective forecasts in opposite directions, forecasts could be more accurate than if only one error or the other were present. This suggests that simply looking at forecast accuracy may hide both of the errors, people may be accidental Bayesians.

*Information Inconsistency.* One of the potential moderators of the advice overweighing is the valence of the advice, and more specifically, the consistency between the advice's valence and consumers' general belief. Confronting inconsistency could be unpleasant. Inconsistency induces negative emotions, and people are aware of such ambivalence, thus take actions to avoid it. On the other hand, as part of human nature, we are in favor of consistent information, for which exerts less cognitive dissonance and consequently requires less cognitive effort to solve the dissonance (Newby-clark, Mcgregor, and Zanna 2002; Olshavsky et al. 1972; Zanna and Lepper 1973). Solving information inconsistency thus is often an automatic process.

Congruity model posits that the evaluation change revises in a direction that eliminates inconsistency (Osgood and Tannenbaum 1955). The solution is to selectively choose one piece of information over the other, or "selective exposure to (consistent) information (Fischer, Schulz-Hardt, and Frey 2008). In this process, information quantity moderates decision-makers' preference for inconsistent vs. consistent post-decision information, that they prefer decision-consistent (vs. inconsistent) information when the

information quantity is high (vs. low) (Yoon, Sarial-Abi, and Gurhan-Canli 2012).

Therefore, combining with the finding of base rate neglect, it is reasonable to expect that consumers would overweigh information that is inconsistent in valence more than that is consistent with their general belief. In other words:

**H5:** The valence inconsistency between base rate belief in a product domain and the reviewer's evaluation moderates the advice overweighing.

### *OVERVIEW OF STUDIES*

To examine how people make affective forecasts based on WOM advice, we first explore the relevance of preference similarity in the real world, examining the predictors of PM in a large archival dataset of Yelp users. Study 1 then examined how people perceive self-other preference similarity (i.e. PM) when evaluating recent movies. Next, Study 2 manipulates PM and Study 3 also manipulated advice valence to examine the effect of PM on how people use advice. Finally, Study 4 replicated Study 3 with a larger sample size of stimuli from a different domain and a more balanced ratio between liked and disliked items of the baseline samples.

### *YELP FIELD STUDY*

To better understand how actual PM information could be displayed and used on real rating websites, we collected field data from Yelp.com to show that 1) we can calculate PM between users in an actual review service, and 2) that PM is a valid cue for preference similarity. Yelp data has been used in the research related to social interaction

thanks to its richness and socially oriented network (Chen and Lurie 2013). Yelp ratings are available for restaurants but also a variety of services such as barbers and dentists (Luca and Zervas 2015). Yelp requires users to create a profile that publically displays the user's first name and last initial, a photo, their location, and other details about the user's background. Although Yelp already allowed users to follow and "friend" other users, they introduced additional functionality in 2010 to allow users to connect with Facebook friends also using Yelp. This function may have been motivated by the expectation that users are more likely to use advice from their friends, perhaps due to their greater perceived PM. Due to this social integration, Yelp provides an ideal data source featuring abundant reviews and a rich social network.

### *Data*

We retrieved data in April 2015 from the "Yelp Dataset Challenge," an annual research competition organized by Yelp. This dataset covers 61,000 businesses located in 10 cities (Pittsburgh, Charlotte, Urbana-Champaign, Phoenix, Las Vegas, Madison, Montreal, Waterloo, Karlsruhe, and Edinburgh) across 4 different countries. The dataset includes reviews and information from 366,000 Yelp users, including each user's ID, number of Yelp reviews, ratings for those reviews, and most importantly their friends' user IDs. These users generated a total of 1.6 million reviews and formed a social network with a total of 2.9 million edges.

### *Results*

By cross-referencing user IDs and business IDs, we created a matrix of matches between friends and non-friends on reviews that were made for the same businesses. For

example, suppose that User A made 20 reviews and among them, User B's also happened to review 5 of the same businesses. Of these 5 common reviews, their ratings agreed on 3 businesses. In the user-pair matrix, we recorded the number of common reviews, 5, the number of matches, 3, and the PM, 60%. This matrix has over 133 billion pairs of users, although the vast majority of user-pairs had zero common reviews, mostly due to the reviewers living in different cities. We thus only consider user-pairs with at least 1 common review, which should mostly limit comparisons to users within the same city. Even then, the dataset still has over 180 million user-pairs.

To calculate PM, we calculated the percentage of shared reviews that agreed for each user-pair. Yelp formally translates star ratings as 1 = "Eek. Methinks not," 2 = "Meh. I've experienced better," 3 = "A-OK," 4 = "Yay! I'm a fan," and 5 = "Woohoo! As good as it gets!" Since the likelihood of exactly matched star ratings is low, we also present results using two different ways of calculating "fuzzy" PM between pairs of users that coded star ratings into fewer categories,  $PM_1$  and  $PM_2$ . To calculate  $PM_1$ , we treated 4 or 5 stars as "like" and 1, 2, or 3 stars as "dislike." To calculate  $PM_2$ , we separated 3-star ratings into a separate category. We present results for all three PM definitions.

Considering only user-pairs with at least one common review, the majority of the users are non-friends. Of the 180 million user-pairs with at least one common review, 449 thousand user-pairs are friends on Yelp. More importantly, pairs of non-friends averaged fewer common reviews ( $M = 1.12$ ,  $SD = .66$ , median = 1, max = 216) than pairs of friends did ( $M = 3.84$ ,  $SD = 7.71$ , median = 1, max = 263). (We do not provide inferential statistics since the large sample size means even small differences are extremely

significant.) That is, friends are more likely to go to and review the same businesses than non-friends.

Although this is consistent with friends having common tastes, the data may be distorted by the fact that 83% of non-friend pairs and 49% of friend pairs with more than zero common reviews have exactly one common review. The distributions are heavily skewed even when we only consider comparisons with at least two common reviews (see Figures 2a and 2b). For user-pairs with at least two common reviews, friends averaged 6.86 ( $SD = 10.24$ , median = 3, max = 263) common reviews and non-friends averaged 2.62 ( $SD = 1.82$ , median = 2, max = 216) common reviews. The number of common reviews between friends, however, is still small, implying that the social network aspect of online WOM is not yet fully mature. However, the number of shared reviews should steadily grow as users write more reviews and as the density of the social network increases.

Figures 3a-d show the distributions of PM for the two fuzzy definitions of PM and for friends and non-friends. Importantly, PM differs between friends and non-friends. For user-pairs with at least two common reviews, the mean level of preference similarity was significantly higher for friends (exact match  $M = 37\%$ , median = 38%;  $PM_1$   $M = 67\%$ , median = 67%;  $PM_2$   $M = 62\%$ , median = 60%) than among non-friends (exact match  $M = 32\%$ , median = 33%;  $PM_1$   $M = 62\%$ , median = 50%;  $PM_2$   $M = 56\%$ , median = 50%).

### *Discussion*

The fact that friends on Yelp have higher PM than non-friends suggests that friends have more similar tastes, which is not that surprising given an abundance of evidence for

homophily—that is, people tend to connect with each other based on certain aspects of similarity, including preference similarity (Mcpherson, Smith-lovin, and Cook 2001). Empirically demonstrating this relationship between friendship and higher PM helps to corroborate the real world validity of PM as an operationalization of preference similarity. PM information on review websites therefore potentially provides a useful cue for making affective forecasts based on user reviews.

Notably, Yelp and other review platforms display a user’s friends’ reviews on the top of each review page if any exist. Since friends’ advice will be more likely to be seen, and friends have higher PM, it is important for both businesses and researchers to know whether consumers can accurately take advice from people with various levels of PM. To study this question in more detail, we turn to three experiments that directly manipulate PM to see how people perceive PM information and whether they can use it to make accurate affective forecasts based on other people’s advice.

### *STUDY 1*

In Study 1, we focus on the first part of the advice-taking process: evaluating PM with the adviser. We test H1 and H2 by manipulating whether PM with an anonymous adviser is objectively high (90%) or low (60%).

#### *Methods*

We recruited 79 participants from Amazon Mechanical Turk to rate recent popular movies. We selected these 70 movies from the top grossing movies released between 2012 and 2014 with IMDb ratings between 4.0 and 7.5 ( $M = 6.3$ ) on IMDb’s 1 to 10

scale, deliberately avoiding universally liked or disliked movies to allow for more preference heterogeneity. The movies were listed on a single page along with corresponding movie posters. Participants rated each movie as “like,” “dislike,” or “not watched yet.”

After rating all 70 movies, participants learned that “a randomly selected IMDb user” had seen and rated all 70 movies and saw a summary of how their ratings matched with the IMDb user on the movies that they had both watched. Specifically, participants read the numbers of movies that they had rated as “like” and as “dislike,” and of those movies, the number that the IMDb user also liked (PM+) or also disliked (PM-). We listed this preference similarity information as both fractions and explicit percentages by multiplying the number of liked movies and number of disliked movies by 60% or 90% and rounding to the nearest movie. This rounding process introduced some variation in the actual PM numbers (e.g., 15 of 17 movies rounds to 88% instead of 90%) but because rounding up and down were about as likely, the average PM+ and PM- were close to 60% and 90%.

After reading these statistics, participants were asked to report their perceived PM for both mutually liking and mutually disliking another random movie (“For another random movie from 2012 to 2014, assuming that you like [dislike] the movie, how likely do you think this IMDb user will also *like* [*dislike*] that movie?”), which we refer to as PPM+ and PPM-, respectively. The goal of this study was to compare PPM+ and PPM- with objective PM+ and PM- levels (H1) and to test how PPM+ and PPM- differ (H2).

## Results

There were 2 participants who did not like any of the 70 movies and 13 who did not dislike any, and therefore were excluded from further analysis. Participants saw on average 29 of the 70 movies ( $M = 29.36$ , median = 28.509,  $SD = 13.43$ ), liking 22 ( $SD = 10.69$ ) and disliking 7 ( $SD = 5.60$ ). A mixed-effects two-way ANOVA of PPM on PM level (between: 60% vs. 90%) and valence (within: like vs. dislike) found significant main effects of PM level ( $F(1, 126) = 9.30$ ,  $p < .01$ ) and valence ( $F(1, 126) = 10.16$ ,  $p < .01$ ), but no interaction ( $F(1, 126) = .04$ ,  $ns$ ). For both PM levels, participants self-reported PPM+ was significantly higher than their PPM- (60%: 64.23% vs. 54.68%,  $t(30) = 2.21$ ,  $p < .05$ ; 90%: 75.24% vs. 66.85%,  $t(32) = 2.31$ ,  $p < .05$ ). That is, participants responded to different levels of preference similarity and, for both high and low similarity, perceived higher preference similarity for liking the same movies than for disliking the same movies, consistent with H1 and H2.

Next, we compared participants' reported preference similarity (PPM+/-) to the actual PM+/- percentages that each participant saw based on the number of mutual likes and dislikes between the participant and the adviser as a percentage of the total movies the participant liked or disliked.

Because we displayed PM information as the fraction of mutually liked (disliked) over total liked (disliked) movies, the actual PM participants saw was not exactly 60% or 90%. The average actual PM+ displayed was 59.39% in the 60% condition (median = 60%,  $SD = 2.87\%$ ; not different from 60%:  $t(30) = -1.19$ ,  $ns$ ), and 90.33% in the 90% condition (median = 90%,  $SD = 2.25\%$ ; not different from 90%:  $t(32) = .85$ ,  $ns$ ). The



average actual PM- displayed was 64.71% in the 60% condition (*median* = 60%, *SD* = 14.62%; greater than 60%:  $t(30) = 1.79, p < .05$ ) and 93.45% in the 90% condition (*median* = 92.00%, *SD* = 6.48%; greater than 90%:  $t(32) = 3.06, p < .01$ ). In the 60% condition, 14 of 31 participants saw PM- greater than 60% (whereas 10 saw a greater PM+ than 60%), and in the 90% condition 22 out of 33 saw actual PM- greater than 90% (compared to 12 for PM+). Although the actual PM- displayed was often greater than the intended 60% or 90%, participants nonetheless underestimated PM- and PPM- was lower than PPM+, offering even stronger support for H2.

Table 1 shows the deviations of PPM from the actual PM displayed as well as absolute deviations. Mixed-effects ANOVAs on PPM deviation found main effects of PM level ( $F(1, 126) = 19.07, p < .001$ ) and valence ( $F(1, 126) = 22.31, p < .001$ ), but no interaction ( $F(1, 126) = .36, ns$ ). Similarly, a mixed-effects ANOVA on absolute PPM deviation found main effects of PM level ( $F(1, 126) = 5.15, p < .05$ ) and valence ( $F(1, 126) = 17.41, p < .001$ ), but no interaction ( $F(1, 126) = .82, ns$ ).

We next examine planned contrasts. In the 60% condition, PPM+ was 4.84% higher than actual PM+ (*median* = 3%, *SD* = 11.71%; equal to 0%:  $t(30) = 2.30, p < .05$ ), whereas PPM- was 10.03% lower than actual PM- (*median* = -6%, *SD* = 22.50%; equal to 0%:  $t(30) = -2.48, p < .01$ ). Average PPM+ deviation was greater than average PPM- deviation ( $t(30) = 3.43, p < .001$ ). In the 90% condition, deviation of PPM+ from PM+ was -15.09% (*median* = -10%, *SD* = 18.27%; equal to 0%:  $t(32) = -4.74, p < .001$ ) and deviation of PPM- from PM- was -26.61% (*median* = -21%, *SD* = 25.07%; equal to 0%:  $t(32) = -6.10, p < .001$ ). Average PPM+ deviation was again greater than average PPM-

deviation ( $t(32) = 3.22, p < .01$ ). Within the same valence, PPM+ deviation was greater in the 90% than in the 60% condition ( $F(1, 62) = 26.60, p < .001$ ), and PPM- deviation was also greater in the 90% condition than in the 60% condition ( $F(1, 62) = 7.71, p < .01$ ).

### *Discussion*

Study 1 found that participants perceived their preference similarity with an anonymous IMDb user to be lower than is objectively true when actual preference similarity was high (90%), but this perception deviation was smaller when actual preference similarity was low (60%) [H1]. Moreover, regardless of high or low actual preference similarity, people thought they were more likely to like the same movies as others (PPM+) than to dislike the same movies (PPM-) [H2].

One possible issue with our results is that some participants rated far fewer movies than others, which may lead to less updating of perceived preference similarity from those participants' priors (i.e., how similar their movies tastes are to IMDb users in general). We therefore tested whether the number of movies participants rated was related to biased perceptions of preference similarity. Neither the deviation between PM+ and PPM+ nor the deviation between PM- and PPM- was correlated with the number of movies participants watched in the 60% condition (PPM+:  $r(34) = -.19, ns$ ; PPM-:  $r(31) = -.07, ns$ ) or in the 90% condition (PPM+:  $r(39) = -.05, ns$ ; PPM-:  $r(31) = .13, ns$ ). Furthermore, the deviation between PM+ and PPM+ was not correlated with the number of movies participants liked in the 60% condition ( $r(27) = .29, ns$ ) or in the 90% condition ( $r(30) = -.07, ns$ ), suggesting that PPM+ deviations were not driven by having too small a sample to draw inferences from. Similarly, the deviation between PM- and

PPM- was also not correlated with the number of movies participants disliked in either the 60% condition ( $r(18) = .00, ns$ ) or 90% condition ( $r(19) = -.35, ns$ ). Therefore, PPM-deviations were affected by sample size neither.

Another aspect of this study that requires further exploration is the fact that we directly provided PM information as statistical information (fractions and percentages) rather than allowing participants to gather this information more naturalistically by experiencing matches or mismatches on a product by product basis. We therefore change the PM presentation format in the next study and added a control condition where no PM information was provided.

## *STUDY 2*

Having found evidence of biased preference similarity perceptions, we next test how people combine their prior preferences and preference similarity with the adviser to make an affective forecast of how much they will like a target stimulus. Study 2 also explored how people perceive PM information in a different domain, comic strips, and how they use PM information obtained from direct experience. By using the comic strip domain, we can guarantee that our participants have seen the same number of previous products for PM estimation. We can also have participants actually experience the target comic in an online study, something that is impossible with most consumer domains.

### *Methods*

We recruited 179 MTurk participants for an experiment about rating comics. We chose Dilbert comics because the comic is among the most widely read comics in the

U.S. For our baseline sample, we selected 10 Dilbert comics from a pretest ( $N = 36$ ) of 20 Dilbert comics based on them having high ratings variance.

After answering demographic questions, all participants self-reported how much they generally like Dilbert comics (on a continuous scale from “dislike a lot” to “like a lot”) and what percentage of Dilbert comics they like and dislike based on past experience, representing their base rate of liking and disliking Dilbert in general. They then viewed and rated the 10 baseline Dilbert comics on a 4-point scale (i.e., “dislike”, “somewhat dislike”, “somewhat like”, and “like”), forcing participants to express a positive or negative preference for each comic.

Participants next saw PM information based on their condition (60%, 90%, and no PM information). Participants in the 60% and 90% PM conditions saw a table summarizing their rating of each comic along with the ratings of the same 10 comics from another “randomly selected participant from a previous study” (the “adviser,” although we never mentioned this term in the study). Specifically, the adviser was assigned to have either a 60% or 90% match in ratings with them. That is, the adviser had the same valence of rating on either 6 or 9 of the same 10 comics, with mismatched ratings for the remainder. Using this table, participants could directly compare their own ratings with the adviser’s ratings, and could thus intuitively estimate their PM. In the control condition, we did not provide any PM information about the randomly selected “adviser.”

Next, participants reported their perceived PM for mutually liking and mutually disliking Dilbert comics, e.g. for mutually liking: “For a random Dilbert comic, assuming

that you like the comic, how likely do you think this previous participant will also like that comic (that is, you both rate the comic either ‘somewhat like’ or ‘like’)?” We then showed that the adviser “liked” the target on an actual rating scale and asked participants to estimate their likelihood of liking and their likelihood of disliking the target comic. Finally, participants answered several questions about their forecasts, the previous participant, their past experience reading Dilbert comics, and the purpose of this study.

### *Results*

Table 2 shows summary statistics for PPM+ and PPM-, self-reported base rates for liking Dilbert comics, forecasts for liking the target comic, and Bayesian forecasts.

*Perceived PM.* A PM condition (control vs. 60% vs. 90%)  $\times$  valence mixed-effects ANOVA on self-reported PPM found main effects of PM condition ( $F(2, 355) = 23.52, p < .001$ ) and PPM+ being higher than PPM- ( $F(1, 356) = 49.90, p < .001$ ), but no significant interaction ( $F(2, 355) = 2.54, ns$ ).

Planned contrasts confirmed that participants perceived higher PPM+ in the 90% PM condition ( $M = 81.24$ ) than in the 60% ( $M = 62.00\%$ ;  $F(1, 176) = 14.97, p < .001$ ) and the control condition ( $M = 66.10\%$ ;  $F(1, 176) = 13.43, p < .001$ ), with no difference in the latter conditions ( $F(1, 176) = .06, ns$ ). Participants also reported higher PPM- in the 90% PM condition ( $M = 64.76\%$ ) than in the 60% ( $M = 53.13\%$ ;  $F(1, 176) = 4.24, p < .05$ ) and the control condition ( $M = 45.56\%$ ;  $F(1, 176) = 23.24, p < .001$ ). PPM- was also lower in the control condition than in the 60% condition ( $F(1, 176) = 7.24, p < .01$ ).

Replicating Study 1, both PPM+ and PPM- were significantly lower than 90% in the 90% PM condition (PPM+  $t(57) = -4.33, p < .0001$ ; PPM-  $t(57) = -6.95, p < .0001$ ).

Unlike in Study 1, PPM+ in the 60% PM condition was not significantly different from 60% ( $t(61) = .97, ns$ ) whereas PPM- was still lower than 60% ( $t(61) = -2.31, p < .05$ ). Participants in the no PM information condition reported an average PPM+ of 66.10%, which is significantly higher than 60% ( $t(58) = 2.32, p < .05$ ), and an average PPM- of 45.56%, which is significantly lower than 60% ( $t(58) = -5.72, p < .0001$ ) and even lower than 50% ( $t(58) = -1.36, p < .05$ ).

*Advice taking.* An ANOVA of participants' forecasts for liking the target comic found a main effect of PM condition ( $F(2, 176) = 3.84, p < .05$ ). Participants forecasted greater likelihood of liking the target comic in the 90% condition ( $M = 80.41\%$ ,  $SD = 21.17\%$ ) than in the 60% ( $F(1, 176) = 4.80, p < .05$ ) and control conditions ( $F(1, 176) = 6.64, p < .05$ ), suggesting that people responded to different levels of PM, consistent with H2. Forecasts were similar in the 60% ( $M = 71.21\%$ ,  $SD = 22.44\%$ ) and the control condition ( $M = 69.46\%$ ,  $SD = 25.18\%$ ;  $F(1, 176) = .18, ns$ ).

We next assess how much these forecasts differed from participants' base rates of liking Dilbert comics—that is, how much participants “updated” their beliefs about the target comic relative to a randomly selected comic based on the WOM they received. The average self-reported base rate for liking Dilbert comics was 60.23% (median = 65%,  $SD = 23.96\%$ , range = 0% to 100%) and the average base rate of disliking Dilbert comics was 33.17% (median = 29%,  $SD = 23.50\%$ , range = 0% to 100%). Because not all participants' liking and disliking base rates added to 100%, we treated the remaining percentage as the probability of a neutral liking. However, since we did not allow neutral ratings in our study, we scaled base rates of liking and disliking to sum to 100%. Doing

so also normalizes those base rates which added up to greater than 100%. After rescaling, the average base rate for liking Dilbert comics was 64.75% (median = 70%,  $SD = 24.25\%$ ). We use these rescaled base rates for all analyses and also scaled the actual forecasts to sum to 100%. Table 2 shows the scaled base rates and forecasts.

Participants' forecasts were significantly higher than their base rates of liking Dilbert comics in all three conditions (*control*:  $M_{BR} = 64.26\%$ ,  $SD = 25.83\%$ ,  $t(58) = 3.00$ ,  $p < .01$ ; *60%*:  $M_{BR} = 62.04\%$ ,  $SD = 26.62\%$ ,  $t(60) = 4.28$ ,  $p < .001$ ; *90%*:  $M_{BR} = 68.09\%$ ,  $SD = 19.52\%$ ;  $t(57) = 4.60$ ,  $p < .001$ ), suggesting that they were influenced by the positive WOM. We next determined how much participants updated from the base rate by calculating the difference between each participant's actual forecast and base rate. The actual forecasts were 12.11% ( $SD = 20.04\%$ ) higher than base rates in the 90% condition, 12.34% ( $SD = 22.50\%$ ) higher in the 60% condition, and 6.81% ( $SD = 17.45\%$ ) higher in the control condition. The degree of updating did not differ across conditions ( $F(2, 175) = 1.43$ , *ns*).

*Forecast Accuracy.* To formally assess the accuracy of advice utilization, we compare participants' predicted likelihoods of liking the target comic with the forecast made by a Bayesian forecaster based on each participant's base rates and the objective PM information (Bayesian<sub>PM</sub>). We calculated *forecast error* by taking calculating the difference between participants' forecasts to Bayesian<sub>PM</sub> (actual forecast - Bayesian<sub>PM</sub>).

A one-way ANOVA of forecast error found a main effect of PM condition ( $F(2, 175) = 235.62$ ,  $p < .001$ ). The forecast error was lower than 0 in the 90% condition ( $M = -13.40\%$ ;  $t(57) = -5.38$ ,  $p < .001$ ), but was greater than 0 in the 60% condition ( $M =$

5.86%,;  $t(60) = 2.09, p < .05$ ). That is, relative to a Bayesian, participants under-forecasted in the 90% condition and over-forecasted in the 60% condition.

*Advice Weighting Accuracy.* To better understand how participants made their forecasts based on the WOM and PM information, we first examined correlations between participants' forecasts and the available information, i.e. their base rates and the perceived PM. In the 60% condition, forecasts were positively correlated with participants' base rates ( $r(59) = .58, p < .001$ ) and with PPM+ ( $r(59) = .26, p < .05$ ), but were not correlated with PPM- ( $r(59) = -.05, ns$ ). In the 90% condition, forecasts were again positively correlated with participants' base rates ( $r(56) = .52, p < .001$ ) and with PPM+ ( $r(56) = .72, p < .001$ ), but were not correlated with PPM- ( $r(56) = -.14, ns$ ). That is, it seems that participants incorporated both base rate and PM information into their forecast.

To assess whether participants accurately weighed WOM advice, we grant participants their subjective PPM+ and PPM- in order to calculate Bayesian<sub>PPM</sub>—what a Bayesian forecaster would forecast based on the participant's subjective PPM. Comparing participants' forecasts to Bayesian<sub>PPM</sub> allows us to assess how accurately participants weigh the advice based on their own perceptions of preference similarity. We thus calculated *weighting error* by subtracting the Bayesian<sub>PPM</sub> from participants' actual forecasts.

The weighting error differed across conditions ( $F(2, 175) = 3.72, p < .05$ ). In the control condition, participants' forecasts ( $M = 71.06\%, SD = 25.03\%$ ) were higher than Bayesian<sub>PPM</sub> ( $M = 66.28\%, SD = 28.72\%$ ;  $M_{error} = 4.78\%, t(57) = 2.19, p < .05$ ). That is,



even people with no information about an adviser's preferences updated their beliefs too much by overweighing the adviser's rating. In the 60% condition, participants' forecasts ( $M = 74.38\%$ ,  $SD = 21.22\%$ ) were also higher than Bayesian<sub>PPM</sub> ( $M = 67.01\%$ ,  $SD = 25.42\%$ ;  $M_{error} = 7.37\%$ ,  $t(61) = 2.69$ ,  $p < .01$ ). However, in the 90% condition, actual forecasts ( $M = 80.21\%$ ,  $SD = 21.38\%$ ) were not significantly different from Bayesian<sub>PPM</sub> ( $M = 82.01\%$ ,  $SD = 19.30\%$ ;  $M_{error} = -1.80\%$ ,  $t(58) = -.76$ , *ns*).

*Moderators.* Finally, we examine how individual differences, such as participants' base rates for liking Dilbert comics, affected forecast accuracy. Indeed, we found that the base rate of liking Dilbert comics was negatively correlated with the weighting error ( $r(176) = -.44$ ,  $p < .001$ ). A linear regression of weighting error on base rate and PM shows that the base rate of liking Dilbert comics predicts the weighting error negatively ( $\beta = -.44$ ,  $t(175) = -6.48$ ,  $p < .001$ ). The less people generally like Dilbert comics, the more they overweighed the advice, which was positive in valence in this study. In other words, people seem to overweigh advice that provides opposite opinion to their base rate, but underweigh advice that is consistent with their base rate.

### *Discussion*

Study 2 found the same PPM patterns as in Study 1: People underestimated 90% PM and perceived PPM+ higher than PPM- (H1 and H2). People did respond to different levels of PM (H3) but they overweighed advice (H4) in the 60% condition. Participants' forecast error was positive in the 90% condition but was negative in the 60% condition. This result suggests that in the 90% condition, given that participants accurately weighted the advice, underestimating PM is what biased their forecasts. In contrast, participants in

the 60% condition overweighed the advice while perceiving PM accurately, leading to a forecast overshooting the Bayesian standard.

Interestingly, we therefore expect that if both the perception error and weighing error appear at the same time with the predicted direction, they should offset with each other and result in an accurate forecast. In this study, we have only found one bias or the other, but our next study finds a condition where both biases occur.

*Positive advice vs. negative advice.* In this study, we only provided positive advice. Rating valence is one of the most influential and essential features of online WOM (Chen, Wang, and Xie 2011; Chevalier and Mayzlin 2006; Liu 2006; Sood et al. 2016). Negative information in particular is more salient and memorable than positive information and thus should have greater impact on WOM (Chen and Lurie 2013; Wu 2013). Furthermore, when the advice valence is the same as what people generally believe, it is plausible that they will not move too much from base rate due to less room left approaching 100%, thus not overweighed the advice. Such limitation will not appear if base rate is not too close to the upper limit.

More importantly, as the analysis supported H5, individual base rate serves as a moderator to the overweighing of advice. It is possible that information that is inconsistent with people's general belief attracts more attention than consistent advice, which could be automatically screened off for less informational value. Therefore, it was negative advice that participants who generally liked products in the domain took into serious consideration and symmetrically, participants who generally disliked products in the domain would weigh more on positive advice.

Because individual base rate in a specific domain is difficult to manipulate, Study 3 manipulated rating valence to test these forecasts. We therefore expect that when making forecasts about disliking a product after seeing negative advice, the weighting error will also be negatively associated with base rates, and that people should overweigh negative advice on average as their base rates indicate generally liking that product domain.

### STUDY 3

#### *Methods*

A total of 273 Amazon Mechanical Turk participants participated in a 3 (PM: 60% vs. 90% vs. no info)  $\times$  2 (rating valence: positive vs. negative) between-subjects design. The study design exactly followed Study 2, with two exceptions. First, to increase the relevance of self-reported base rates, we asked for general liking and disliking of Dilbert comics immediately *after* participants rated the 10 baseline stimuli. Second, the previous participant's rating of the 11<sup>th</sup> comic was randomly assigned to be either 5 stars (positive advice valence) or 1 star (negative).

#### *Results*

*Perceived Preference Match.* A PM condition  $\times$  valence mixed effects ANOVA on self-reported PPM found main effects of PM condition ( $F(2, 543) = 21.33, p < .001$ ) and valence ( $F(1, 544) = 94.97, p < .001$ ), and an interaction effect of PM condition  $\times$  valence ( $F(2, 543) = 9.08, p < .001$ ). Planned contrasts showed that PPM+ was higher than PPM- in all three conditions, again supporting H2 (control: PPM+  $M = 74.09\%$ , PPM-  $M = 48.55\%$ ,  $F(1, 270) = 58.05, p < .001$ ; 60%: PPM+  $M = 62.10\%$ , PPM-  $M = 54.88\%$ ,  $F(1,$

270) = 5.06,  $p < .05$ ; 90%: PPM+  $M = 83.72\%$ , PPM-  $M = 61.42\%$ ,  $F(1, 270) = 47.84$ ,  $p < .001$ ).

Participants reported higher PPM+ in the 90% condition than in the 60% condition ( $F(1, 185) = 87.36$ ,  $p < .001$ ). However, participants reported the same PPM- in the 90% condition as in the 60% condition ( $F(1, 185) = 2.38$ , *ns*), with both being higher than in the control condition ( $F(1, 177) = 9.32$ ,  $p < .01$ ;  $F(1, 178) = 3.37$ ,  $p = .07$ ). Replicating Studies 1 and 2, both PPM+ and PPM- were much lower than 90% in the 90% condition (PPM+  $t(92) = -4.16$ ,  $p < .001$ ; PPM-  $t(92) = -8.34$ ,  $p < .001$ ), indicating participants greatly underestimated high PM level.

In addition, we worked on two sets of base rates, Explicit Base Rate that participants self-reported and Implicit Base Rate that is calculated from their 10 actual ratings. By asking about Explicit Base Rates after rating 10 Dilbert comics, the concern of selection bias of our stimuli and that participants may further update their base rate in the Explicit Base Rate questions after seeing 10 stimuli was strictly controlled by that very little distance of the Implicit Base Rate calculated from their actual ratings (for positive ratings  $M = 78.35\%$ ,  $SD = 21.72\%$ ; for negative ratings  $M = 21.65\%$ ,  $SD = 21.72\%$ ) had to the self-reported Explicit Base Rate (for positive ratings  $M = 70.51\%$ ,  $SD = 23.52\%$ ; for negative ratings  $M = 27.19\%$ ,  $SD = 23.47\%$ ). However, to be consistent with Study 2, we looked into the Explicit Base Rate and used it to calculate the normative forecast to judge the weighting and forecast accuracy.

Again, to cope with the binary scale that we used for calculating Bayesian predications, we scaled the base rates to 100%, omitting the neutral ratings or calibrating

the base rates which added up to greater than 100%. The scaled Base rates of likes and dislikes are also shown in Table 3.

The factor scores from our six questions on the adviser (asking for similarity, trustworthiness, etc.) ( $\beta = .32, t(268) = 5.47, p < .001$ ) with the base rate of likes ( $\beta = .39, t(268) = 5.55, p < .001$ ) and how much they liked Dilbert relative to others (liked more or less compared to others) ( $\beta = -.17, t(273) = -2.43, p < .05$ ) together predicted PPM+ out of other variables (BR- and the set PM), suggesting that people perceived preference matching for positive ratings at least partially from their internal belief on how similar other people's preferences are to their own. However, the same regression analysis on PPM- shows that the set PM ( $\beta = .18, t(268) = 2.96, p < .01$ ) and the base rate of likes ( $\beta = -.25, t(268) = -3.28, p < .001$ ) together predicted PPM-. How the adviser's negative ratings matching with participants' negative ratings in the past and their base rates had the most powerful impact on the perception of PM-, rather than other factors.

*Advice taking.* A linear regression analysis of three items (forecast, base rate of likes, and PPM) the same as Study 2 shows that the forecast of liking (scaled) was based on both the BR+ ( $\beta = .64, t(141) = 10.82, p < .001$ ) and the PPM+ ( $\beta = .24, t(141) = 4.06, p < .001$ ) in the positive rating condition. But the forecast of disliking was based only on the Base Rate of dislikes ( $\beta = .55, t(126) = 7.41, p < .001$ ). Therefore, in this study, participants considered the preference matching with the adviser as the diagnostic information only when making the positive forecast (and positive advice was provided), suggesting that correct information was used. But participants largely depended on the base rate of dislikes in the forecast of future disliking.

*Forecast Accuracy.* The actual forecast using *positive* advice was not different from the normative forecast (Bayesian<sub>PM</sub>) in the 60% condition (Forecast:  $M = 75.86\%$ , median =  $80.57\%$ ,  $SD = 22.96\%$  vs. Bayesian<sub>PM</sub>:  $M = 73.25\%$ , median =  $83.86\%$ ,  $SD = 27.23\%$ ;  $t(43) = 1.03$ , *ns*), but was much lower than the normative forecast in the 90% condition (Forecast:  $M = 85.95\%$ , median =  $90\%$ ,  $SD = 13.62\%$  vs. Bayesian<sub>PM</sub>:  $M = 95.75\%$ , median =  $97.30\%$ ,  $SD = 4.34\%$ ;  $t(50) = -5.25$ ,  $p < .001$ ). When using *negative* advice, participants forecasts higher than the normative forecast (Bayesian<sub>PM</sub>) in the 60% condition (Forecast:  $M = 49.06\%$ , median =  $50.00\%$ ,  $SD = 30.34\%$  vs. Bayesian<sub>PM</sub>:  $M = 35.90\%$ , median =  $30.30\%$ ,  $SD = 26.64\%$ ;  $t(49) = 4.29$ ,  $p < .001$ ), but forecasted quite accurately in the 90% condition (Forecast:  $M = 64.36\%$ , median =  $74.81\%$ ,  $SD = 26.94\%$  vs. Bayesian<sub>PM</sub>:  $M = 66.17\%$ , median =  $69.53\%$ ,  $SD = 22.55\%$ ;  $t(41) = -.45$ , *ns*).

The forecast error calculated by subtracting the actual forecast from the Bayesian<sub>PM</sub> showed the same results as the descriptive data. Forecast error when using *positive* advice was not different from 0 in the 60% condition ( $M = 2.60\%$ , median =  $0\%$ ,  $SD = 16.85\%$ ;  $t(44) = 1.03$ , *ns*), but was smaller than 0 in the 90% condition ( $M = -9.80\%$ , median =  $-5.45\%$ ,  $SD = 13.33\%$ ;  $t(50) = -5.25$ ,  $p < .001$ ). On the other hand, forecast error when using *negative* advice was greater than 0 in the 60% condition ( $M = 13.16\%$ , median =  $7.68\%$ ,  $SD = 21.69\%$ ;  $t(49) = 4.29$ ,  $p < .001$ ), but was not different from 0 in the 90% condition ( $M = -.28\%$ , median =  $2.5\%$ ,  $SD = 27.21$ ;  $t(41) = -.45$ , *ns*).

*Advice Weighting Accuracy.* The weighting error using *positive* advice is consistent with the finding in Study 2, that participants accurately assigned weights to the advice in the 90% condition (Bayesian<sub>PPM</sub>:  $M = 88.55\%$ , median =  $91.76\%$ ,  $SD = 12.52\%$ ), and in

the 60% condition ( $M = 74.69\%$ , median = 80.91%,  $SD = 27.01\%$ ). When using *negative* advice, participants largely *over-weighed* the negative advice in both the 60% condition by comparing the actual forecast and the Bayesian<sub>PPM</sub> ( $M = 34.80\%$ , median = 29.64%,  $SD = 29.35\%$ ) and in the 90% condition ( $M = 50.06\%$ , median = 50.84%,  $SD = 27.91\%$ ).

In addition to descriptively comparing the actual forecast and the Bayesian<sub>PPM</sub>, similarly to Study 2, we also analyzed the weighting error, by subtracting the Bayesian<sub>PPM</sub> from the actual forecast. Consistent with the descriptive data, the weighting using *positive* advice was accurate in the 90% condition (weighting error:  $M = -2.60\%$ , median = -2.50%,  $SD = 11.62\%$ ; not different from 0:  $t(50) = -1.60$ , *ns*) and in the 60% condition (weighting error:  $M = 1.17\%$ , median = -.50%,  $SD = 17.35\%$ ; not different from 0:  $t(43) = .45$ , *ns*). However, participants overweighed *negative* advice in both the 60% (weighting error:  $M = 14.26\%$ , median = 9.12%,  $SD = 23.86\%$ ; greater than 0:  $t(49) = 4.23$ ,  $p < .001$ ) and the 90% condition (weighting error:  $M = 14.30\%$ , median = 10.65%,  $SD = 29.76\%$ ; greater than 0:  $t(41) = 3.11$ ,  $p < .01$ ). Notably, the weighting using positive advice in the control condition was also accurate (weighting error:  $M = 1.46\%$ , median = .42%,  $SD = 13.51\%$ ; not different from 0:  $t(48) = .76$ , *ns*), and participants overweighed negative advice in the control condition (weighting error:  $M = 9.05\%$ , median = 5.03%,  $SD = 27.06\%$ ; greater than 0:  $t(35) = 2.01$ ,  $p < .05$ ) like in the other conditions.

*Moderators.* We further analyzed on individual level forecasts, and found that when the advice was positive, the base rate of liking Dilbert comics was negatively correlated with the weighting error ( $r(142) = -.33$ ,  $p < .001$ ). After controlling for the PM level, the

base rate of liking Dilbert comics predicts the weighting error ( $\beta = -.33, t(141) = -4.14, p < .001$ ). Similarly, when the advice was negative, the base rate of disliking Dilbert comics was negatively correlated with the weighting error ( $r(127) = -.26, p < .01$ ). After controlling for the PM level, the base rate of disliking Dilbert comics predicts the weighting error ( $\beta = -.26, t(125) = -3.05, p < .01$ ).

More interestingly, we looked into the correlation between PPM bias (PPM – PM) and weighting error to predict the ultimate forecast error. PPM+ bias was not correlated with weighting error in the 60% condition ( $r(42) = -.26, ns$ ) or in the 90% condition ( $r(49) = .10, ns$ ), but PPM- bias was negatively correlated with weighting error in the 60% condition ( $r(48) = -.40, t < .01$ ) and was slightly more negatively correlated with weighting bias in the 90% condition ( $r(40) = -.52, t < .001$ ). Given that mathematically the ultimate forecast error is determined by both PPM bias and weighting bias, therefore, we were not surprising to find that the ultimate forecast in the 90% condition was relatively accurate when the forecast was based on negative advice, given that PPM- bias pointed to the opposing direction to the weighting error in that condition.

### *Discussion*

We demonstrated that the accurate forecast in the 90% condition when using negative advice is mainly an outcome of two errors in opposing directions offsetting each other, i.e. a downgrading effect of perception a lower PM than the given with an upward force of overweighing negative advice. That was confirmed by the negative correlation between PPM- bias and weighting bias. In taking negative advice, people put higher weight than should on advice from someone they do not consider as similar with them as



the actual. This supports H4 that people overweigh inconsistent information and also the negative information, regardless of the perceived preference similarity.

In this study, we showed that the quality of Bayesian inference about future liking/disliking is moderated by the advice valence [H5]. Positive and consistent advice was less weighed than should, whereas negative and inconsistent advice was overweighed compared with the Bayesian standard. Accidentally, people are able to make accurate Bayesian inference when taking an inconsistent advice from a similar other, although they are not able to truly think in a Bayesian way and even realize the adviser is quite similar to them.

#### *STUDY 4*

In previous studies, we manipulated PM based on the total number of matched ratings, regardless of the valence. To rule out the possibility that the difference between PPM+ and PPM- is due to the difference between the actual PM+ and the actual PM-, we conducted Study 4 in a different domain, movies, while directly calculating and displaying PM+ and PM- for participants to reduce biases due to memory errors while sampling. Besides, the participants would experience fewer liked items, so that the ratio between the positive and the negative rating matches was smaller than previous studies. We aimed to replicate our findings in previous studies after measuring the PM perception accuracy and forecast accuracy based on a bigger sample size and a different domain.

## *Methods*

We recruited 398 Amazon Mechanical Turk participants for a 3 (between-subjects: PM 60% vs. 90% vs. no info)  $\times$  2 (within-subjects advice order: negative vs. positive first) mixed design. After participants were asked about the average number of movies they watch every month and were screened on watching at least two movies a month on average, they rated between 20 to 25 movies from the 70 movies in Study 1 (in predetermined order) as “liked,” “disliked,” or “not watched.” Participants finished the rating task as soon as they rated at least 10 movies as “liked” and at least 10 movies as “disliked,” or when they rated a total of 25 watched movies. A few participants had not watched at least 20 of the 70 movies and therefore rated fewer than 20 total movies (97 out of 398 participants).

As in Study 3, we asked for general liking and disliking of popular movies immediately *after* participants rated movies. Participants in the 60% and the 90% PM conditions then saw summary statistics for the number of movies that they had liked and disliked, and the number of movies that the previous IMDb user liked and disliked in common. These numbers were rounded after multiplying the number of liked and disliked movies by 60% or 90%. We also displayed a message that the approximate match rate was “about 60% (90%),” depending on condition. No statistics were provided for the control group. Following the PM information page, participants reported their PPM for both mutual liking and disliking “a random new movie released this year”.

For the forecasting task, participants forecasted how likely they would like and dislike two additional newly released movies. All participants saw a positive rating and a

negative rating from “this randomly selected IMDb users”. These two differently valenced ratings were displayed in a random order, so that the participants would see the previous IMDb user’s rating for the first movie to be either positive or negative, and so was for the second movie. Finally, they viewed the poster and a one-line description of the two yet-to-be-released target movies, and reported how interested they were in watching each movie. The study ended with several questions about the study purpose and their past experience with similar studies if any.

### *Results*

On average, participants rated 21 movies (median = 25,  $SD = 5.88$ ), liking 15 (median = 16,  $SD = 5.82$ ), and disliking 6 (median = 5,  $SD = 4.57$ ). The ratio between the number of movies liked and the number of movies disliked was better controlled in this study (15:6) than in Study 1 (21:6). Because in this study we set a maximum number of sample movies, participants may vary in the total number of movies rated. We found that the number of movies they watch every month significantly predicts the total number of movies rated in this study ( $\beta = .18$ ,  $t(795) = 5.30$ ,  $p < .001$ ). More importantly, how frequently the participants watch movie also significantly predicts the base rate ( $\beta = .38$ ,  $t(795) = 11.66$ ,  $p < .001$ ). In the following analysis, therefore, we will control for the effect of total number of movies rated.

*Perceived Preference Match.* For PPM+, there was a main effect of the given PM ( $F(2, 395) = 30.06$ ,  $p < .001$ ). A further contrast analysis shows that participants reported PPM+ of 66.09% (median = 70%,  $SD = 17.12\%$ ) in the control condition, not different from the PPM+ of 65.06% in the 60% condition (median = 65%,  $SD = 14.91\%$ ,  $F(1, 395)$

= .26, *ns*). But the PPM+ was much higher in the 90% condition ( $M = 79.19\%$ , median = 85%,  $SD = 17.19\%$ ) than in the 60% condition ( $F(1, 395) = 49.11, p < .001$ ).

The PPM- ( $M = 57.41\%$ , median = 60%,  $SD = 24.25\%$ ), on the other hand, was much lower than the PPM+ overall ( $M = 70.16\%$ , median = 71%,  $SD = 17.63\%$ ;  $t(397) = 10.70, p < .001$ ), and in the control condition ( $t(131) = 9.88, p < .001$ ). An overall ANOVA shows a main effect of PM level on PPM- ( $F(2, 395) = 34.14, p < .001$ ). PPM- in the control condition ( $M = 45.77\%$ , median = 43%,  $SD = 21.18\%$ ) was lower than that in the 60% condition ( $M = 57.79\%$ , median = 60%,  $SD = 20.18\%$ ;  $F(1, 395) = 18.93, p < .001$ ), which was also lower than that in the 90% condition ( $M = 68.51\%$ , median = 80%,  $SD = 25.15\%$ ;  $F(1, 395) = 15.16, p < .001$ ). Therefore, with fewer liked movies and thus better ratio between the liked and the disliked, we replicated our finding in Study 1, and further demonstrated that the default level of PPM for disliked items is different from that for liked items.

But given that the number of movies participants potentially rated ranged from as small as 1 to maximum 25, we need a further test of the individual sample size on the perceived PM. This analysis is particularly important to explore whether our findings of PPM are subject to the sample size. We split the number of liked movies and disliked movies from their median, and created two sets of dummy variables of sample size for liked and disliked movies (dummy variable: larger than the median vs. smaller or equal to the median). We found that both the PM level ( $F(2, 392) = 31.12, p < .001$ ) and the dummy variable of liked movies ( $F(1, 392) = 9.77, p < .01$ ) including their interaction term ( $F(2, 392) = 3.34, p < .05$ ) had an effect on PPM+ overall. However, a following

detailed analysis only shows an effect of sample size on PPM+ in the 60% condition ( $F(1, 130) = .09, ns$ ).

For PPM-, only the PM level had an effect ( $F(2, 392) = 33.50, p < .001$ ). It was only in the 90% condition that the sample size of disliked movies had an effect on PPM- ( $F(1, 132) = 4.49, p < .05$ ), but not in the control condition ( $F(1, 130) = .01, ns$ ), nor in the 60% condition ( $F(1, 130) = .09, ns$ ).

Therefore, we demonstrated that our finding, although sometimes affected by sample size, should at least partially hold valid that PM level largely affects the perception of PM.

*Advice taking.* The same linear regression analysis of three items (forecast, base rate of likes, and PPM) as in Study 2 and 3 shows that after controlling for the order effect (no effect:  $\beta = 2.01, ns$ ), the forecast of liking (scaled) was based on both the BR+ ( $\beta = .49, p < .001$ ) and the PPM+ ( $\beta = .51, p < .001$ ) in the positive rating condition. However, inconsistent with Study 3, the forecast of disliking was based only on the PPM- ( $\beta = .21, p < .01$ ) rather than Base Rate of dislikes. Again, participants used correct information when making the positive forecast (and positive advice was provided), but largely depended on PPM- in the forecast of future disliking.

*Forecast Accuracy.* We measured the forecast accuracy by subtracting the Bayesian<sub>PM</sub> from the actual forecast. But different from previous studies, since we actually displayed PM information in statistics, we calculated Bayesian<sub>PM</sub> using the displayed PM, rather than the set PM, for the purpose of accuracy. As this study had a within-subject design feature, we first reported the forecast accuracy of the first order

movie with positive and negative advice respectively, and then analyzed the whole dataset for the main effect and possible interactions, counting the within-subject factor.

When the advice is *positive*, participants made a forecast for the first order movie accurately in the 60% condition ( $M = -1.59\%$ , median =  $-1.58\%$ ,  $SD = 17.68\%$ ; not different from 0:  $t(65) = -.73$ , *ns*), but much lower than the normative in the 90% condition ( $M = -14.28\%$ , median =  $-10.18\%$ ,  $SD = 13.07\%$ ; smaller than 0:  $t(63) = -8.74$ ,  $p < .001$ ). But when the advice is *negative*, participants made a forecast higher than the normative in the 60% condition ( $M = 11.95\%$ , median =  $10.76\%$ ,  $SD = 25.55\%$ ; greater than 0:  $t(65) = 3.80$ ,  $p < .001$ ), but lower than the normative in the 90% condition ( $M = -12.59\%$ , median =  $-15.43\%$ ,  $SD = 24.62\%$ ; smaller than 0:  $t(69) = -4.28$ ,  $p < .001$ ).

An analysis on the three-way interaction with one within-subject factor (forecast order) and two between-subject factors (valence  $\times$  PM level) shows that there was no effect of valence or interaction terms or an order effect, but only the PM level ( $\beta = 38.37$ ,  $\chi^2(1, N = 532) = 102.51$ ,  $p < .01$ ) that had an effect on the forecast error.

*Advice Weighting Accuracy.* In this study, we also measured the weighting error as another dependent variable by subtracting the Bayesian<sub>PPM</sub> from the actual forecast. It was calculated to measure how accurate participants assigned weight on advice when making a forecast about their future liking and disliking. Again, we normalized the base rates ( $BR_{\text{explicit}}$ ) to 100%. Similar to the forecast error analysis, we will first report the weighting error using positive and negative advice respectively for the first order movie, and then examine the possible interaction terms and order effect counting the within-subject factor and using the full dataset.

Different from Study 2 and 3, participants slightly underweighed the *positive* advice of the first order movie in the 90% condition ( $M = -3.56\%$ , median =  $-3.98\%$ ,  $SD = 15.46\%$ ; smaller than 0:  $t(63) = -1.84$ ,  $p < .05$ ). The weighting in the 60% condition was also lower than the normative ( $M = -4.19\%$ , median =  $-4.49\%$ ,  $SD = 18.94\%$ ; smaller than 0:  $t(65) = -1.80$ ,  $p < .05$ ), but the weighting in the control condition was accurate ( $M = 1.08\%$ , median =  $2.54\%$ ,  $SD = 20.53\%$ ; not different from 0:  $t(66) = .43$ ,  $ns$ ).

Also different from Study 2 and 3, participants accurately weighed the *negative* advice of the first order movie in the 90% condition ( $M = 3.13\%$ , median =  $-.23\%$ ,  $SD = 24.79\%$ ; not different from 0:  $t(69) = 1.06$ ,  $ns$ ), although overweighed in the 60% condition ( $M = 9.92\%$ , median =  $6.31\%$ ,  $SD = 26.19\%$ ; greater than 0:  $t(65) = 3.08$ ,  $p < .01$ ), and control condition ( $M = 12.07\%$ , median =  $8.35\%$ ,  $SD = 28.04\%$ ; greater than 0:  $t(64) = 3.47$ ,  $p < .001$ ).

An analysis on the three-way interaction with one within-subject factor (valence  $\times$  PM level  $\times$  forecast order) shows that only the valence ( $\beta = -9.08$ ,  $\chi^2(1, N = 796) = 29.24$ ,  $p < .01$ ) and the PM level ( $\beta = -4.40$ ,  $\chi^2(2, N = 796) = 6.38$ ,  $p < .05$ ) had an effect on the weighting error, but with no order effect nor interaction effects.

A further two-way ANOVA shows that there was no interaction effect of valence  $\times$  PM level ( $F(2, 790) = .82$ ,  $ns$ ), but only main effects of the valence ( $F(1, 790) = 28.93$ ,  $p < .001$ ), and the PM level ( $F(2, 790) = 3.18$ ,  $p < .05$ ) on the weighting error. A contrast analysis shows that the in the 60% condition, participants overweighed the advice ( $M = 5.21\%$ , median =  $3.57\%$ ,  $SD = 24.39\%$ ), marginally higher than in the 90% condition ( $M = 1.32\%$ , median =  $-1.57\%$ ,  $SD = 23.04\%$ ;  $F(1, 790) = 3.58$ ,  $p = .06$ ), but not different

from the control condition ( $M = 6.23\%$ , median = 5.31%,  $SD = 24.85\%$ ;  $F(1, 790) = .24$ , *ns*).

Since there was no order effect, we tested the weighting error with the full dataset, and found that participants accurately weighed *positive* advice in all PM conditions, but overweighed *negative* advice in all conditions (the control condition:  $M = 11.93\%$ , median = 10.19%,  $SD = 27.68\%$ , greater than 0:  $t(131) = 4.95$ ,  $p < .001$ ; the 60% condition:  $M = 9.95\%$ , median = 7.73%,  $SD = 27.05\%$ , greater than 0:  $t(131) = 4.23$ ,  $p < .001$ ; the 90% condition:  $M = 4.43\%$ , median = 4.42%,  $SD = 25.49\%$ , greater than 0:  $t(133) = 2.01$ ,  $p < .05$ ).

*Moderators.* Considering the fact that individual participant may significantly vary in base rate of liking and disliking movies, besides of analyzing on aggregate level, we also tested the effect of individual base rate level on weighting error. We found that weighting error was negatively correlated with base rate of liking movies ( $r(396) = -.56$ ,  $p < .001$ ) when the advice was positive, and was also negatively correlated with base rate of disliking movies ( $r(396) = -.52$ ,  $p < .001$ ) when the advice was negative. After controlling for PM level and total number of movies rated, the base rate (liking movies) significantly and solely predicts the weighting error ( $\beta = -.57$ ,  $t(394) = -12.66$ ,  $p < .001$ ) when the advice was positive, and the base rate (disliking movies) also significantly predicts (together with other two factors) the weighting error ( $\beta = -.56$ ,  $t(394) = -12.28$ ,  $p < .001$ ) when the advice was negative. Therefore, supporting our hypothesis on inconsistent information, people overweigh disparate advice and underweigh consistent advice.



Given our findings on averaging level, we also tested whether people simultaneously made two systematic errors of PM perception and advice overweighting on individual level. We found that the bias term of PM perception (PPM+ – PM+) was negatively correlated with weighting error ( $r(396) = -.21, p < .001$ ) when the advice was positive, and negative PM perception bias (PPM- - PM-) was also negatively correlated with weighting error ( $r(396) = -.31, p < .001$ ) when the advice was negative.

### *Discussion*

We used a different stimulus in this study to test how the perceived PM and the later forecast were affected by a larger sample size with a more balanced number of liked and disliked items of the baseline stimuli. Our previous findings of PPM were confirmed again in this study. However, due to multiple changes in experimental design, e.g. the rating scale, PM information display, and the stimulus, on average participants overweighed the negative advice, but not in exactly the same magnitude, thus making a less accurate forecast. But more importantly, we found again that people made two systematic errors in offsetting directions in perceiving PM level and weighing on advice.

## *GENERAL DISCUSSION*

The results from these four studies suggest that consumers make two errors compared to a Bayesian advice taker. First, they assume low preference similarity (around 60%) with others as the default, thus underestimating preference similarity for similar advisers. Second, and somewhat contrary to the first error, they nonetheless overweighed the adviser's rating and underweighed their own base rates.

Moreover, both of these errors were exacerbated for negative advice. Consumers assume an even lower level of preference similarity (around 50%) for mutually disliked products and put even more weight on negative-valenced advice, to the point of total base rate neglect. This result confirms “base rate neglect” in advice-taking context, contrary to past research not finding base rate neglect in contexts where base rates are endogenous (Koehler 1996).

In the negative advice, one of the other possible explanations for more extreme overweighing is the consistency between advice valence and what people generally believe. Research on information processing suggests that people process preference-inconsistent information differently from preference-consistent information (confirmatory bias: e.g., Kleck and Wheaton 1967; Lord, Ross, and Lepper 1979; Schulz-Hardt et al. 2000; Chernev 2001; cf., Freedman 1965; Trope and Bassok 1982). A major part of the research shows that inconsistent information could be under-evaluated and less recalled than consistent information, referred as “confirmatory bias” (Kleck and Wheaton 1967), but still many others found opposite preference for inconsistent information. Fischer, Schulz-Hardt, and Frey (2008) solves these contradictory theories by suggesting that the preference for consistent versus inconsistent information is moderated by information quantity, such that participants who choose 1 piece from 2 pieces (vs. 10 pieces) of information strongly prefer inconsistent (vs. consistent) information, because different “salient selection criteria” (information direction vs. information quality) are triggered. Since our studies only provide a single source of advice, we expect participants to be more affected by preference-inconsistent information (in this case, negatively valenced

advice because averagely participants generally like the domains) than preference-consistent information.

More pertinent to consumer research, this study addressed some key questions of how consumers utilize WOM advice by adopting a Bayesian framework to assess the whole process of advice taking, from perception of preference similarity to assigning weight on advice. In doing so, we are able to compare advice taking to a normative standard and assess ex ante forecasting errors. Therefore, we answered questions about how people use WOM information and provided insights to the previous disputes about whether people take advice from similar versus dissimilar others. Our efforts to quantify the assessment of advice-taking quality make it possible to disentangle these two opposing errors in the decision-making process, which could lead to accidentally accurate forecasts. Thus, we contribute to the literature of advice-taking a new paradigm that not only examines the psychology when taking other's advice but also focuses on the cognitive limits of the human brain. Our findings have implications for socially-connected online shopping platforms and can help managers better understand consumers who rely on online WOM more than ever before.

#### *Future Directions*

Although we found that consumers underweighed their own base rates for liking of comics, recent works suggest that reliance on base rates may be domain specific. That is, participants may overweigh their past experiences in forecasting their future liking of experiential goods (movies, music, foods etc.) than material goods, given that people rely less on reviews of experiential goods (Dai, Chan, and Mogilner 2015). Future research

should test this distinction more formally, since our study has examined only experiential goods. In general, broadening the domains studied to include more important decisions such as medical decisions and voting behavior seems wide open for exploration.

Another fruitful area for further exploration is to better understand how consumers make preference similarity inferences about others based on less explicit PM information, such as demographics. In a preliminary study, we tested how perceived PM is affected by demographic similarity such as including gender, age, and level of education and surprisingly found that people were not more influenced by advice from demographically similar advisers. Further research is needed to better understand these relationships, including varying domains and looking at demographic dissimilarities. A related question is how preference similarity in one domain carries over to preference similarity in other domains. Two professors may know their shared tastes in research topics but what inferences will they then draw about their shared tastes in music or beer?

Finally, a natural follow-up question regards another important step in the advice-taking process: how does preference similarity affect advice *seeking*? Meng, Chen, and Bartels (2016) provide one take on this, suggesting that people seek advice from similar and dissimilar advisers depending on the type of judgment being made. Our own preliminary studies suggest that there is a strong recency effect in advice taking: people seek advice from those that recently gave them “good” advice (i.e. advice that was proved to be consistent with their actual experience) much more so than one would expect given a larger sample of preference matching information.

## *Conclusion*

In summary, we have taken the first steps in exploring how preference similarity affects how consumers use WOM advice. Our research suggests extensive managerial implications to corporations who deal with massive online WOM that providing preference matching history may facilitate more acceptance of online reviews, as any PM higher than default from a stranger should elicit more usage of advice. Our Bayesian framework could also help businesses better understand the advice-taking process of their consumers, and therefore provides insights to them in an effort to help consumers make better decisions using online reviews. Websites such as Yelp will be able to calculate the Bayesian forecast for user pairs with sufficient shared reviews and aggregating across many users; a side benefit to doing so is that users would be motivated to rate more businesses in order to improve the quality of their own forecasts.

*FIGURES AND TABLES*

FIGURE 1. EXPERIMENTAL PARADIGM USED ACROSS ALL STUDIES.

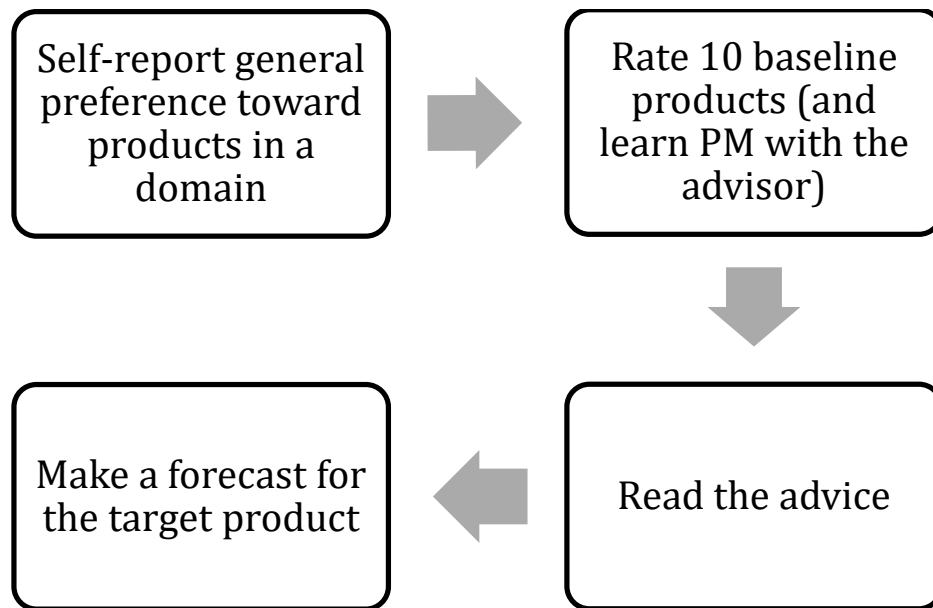


FIGURE 2A. DISTRIBUTIONS OF NUMBER OF COMMON REVIEWS BETWEEN NON-FRIENDS (WITH AT LEAST 2 COMMON REVIEWS).

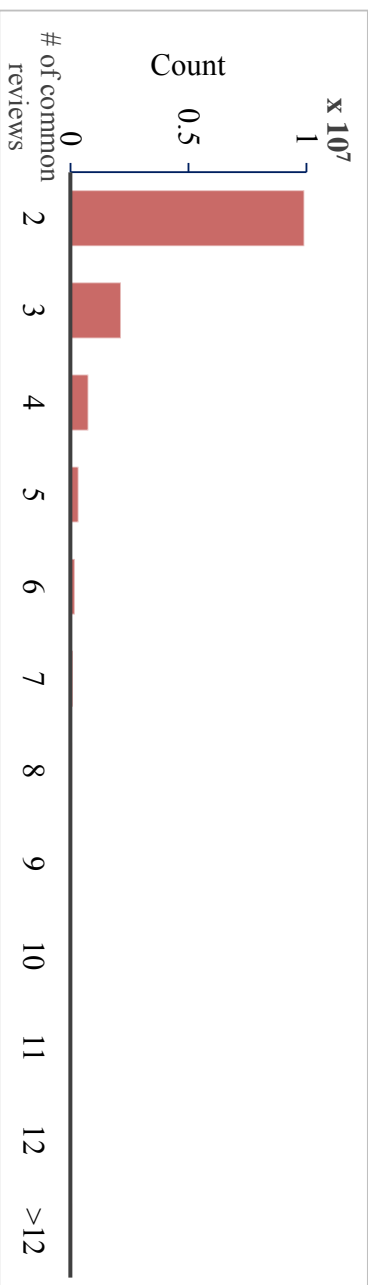


FIGURE 2B. DISTRIBUTIONS OF COMMON REVIEWS BETWEEN FRIENDS (WITH AT LEAST 2 COMMON REVIEWS).

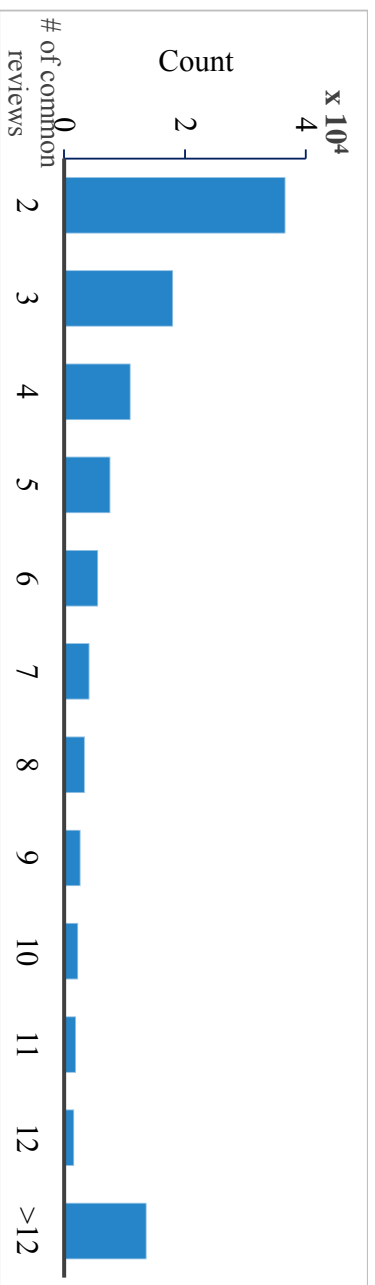


FIGURE 3A. DISTRIBUTIONS OF FUZZY MATCHING PM<sub>i</sub> AMONG NON-FRIENDS (WITH AT LEAST 2 COMMON REVIEWS).

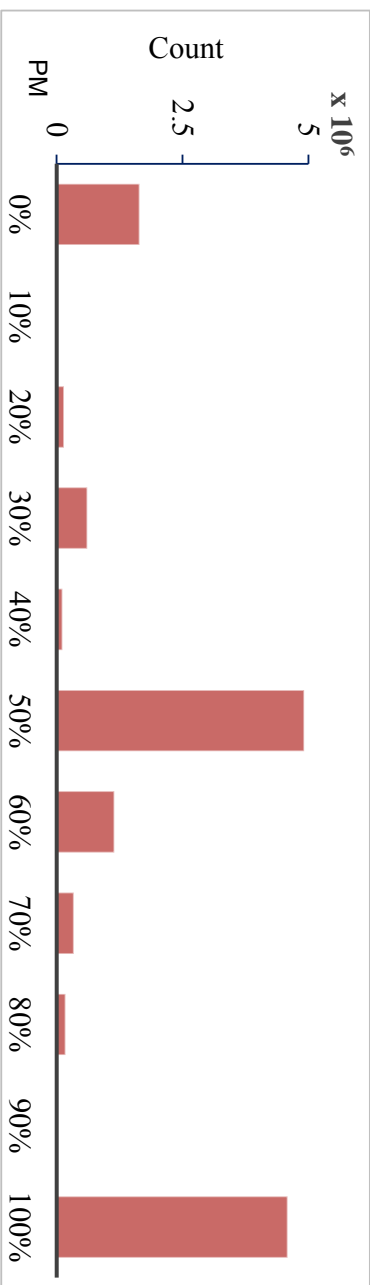


FIGURE 3B. DISTRIBUTIONS OF FUZZY MATCHING PM<sub>i</sub> AMONG FRIENDS (WITH AT LEAST 2 COMMON REVIEWS).

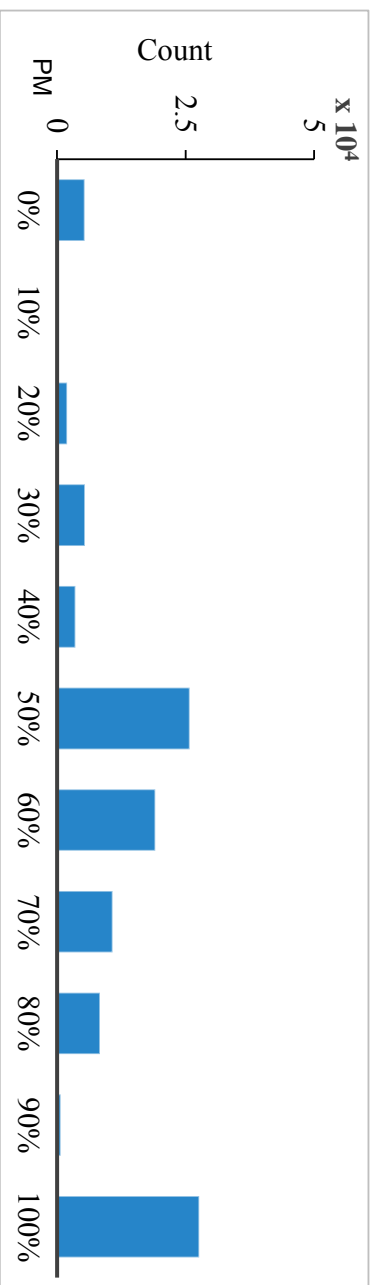




FIGURE 3C. DISTRIBUTIONS OF FUZZY MATCHING PM<sub>2</sub> AMONG NON-FRIENDS (WITH AT LEAST 2 COMMON REVIEWS).

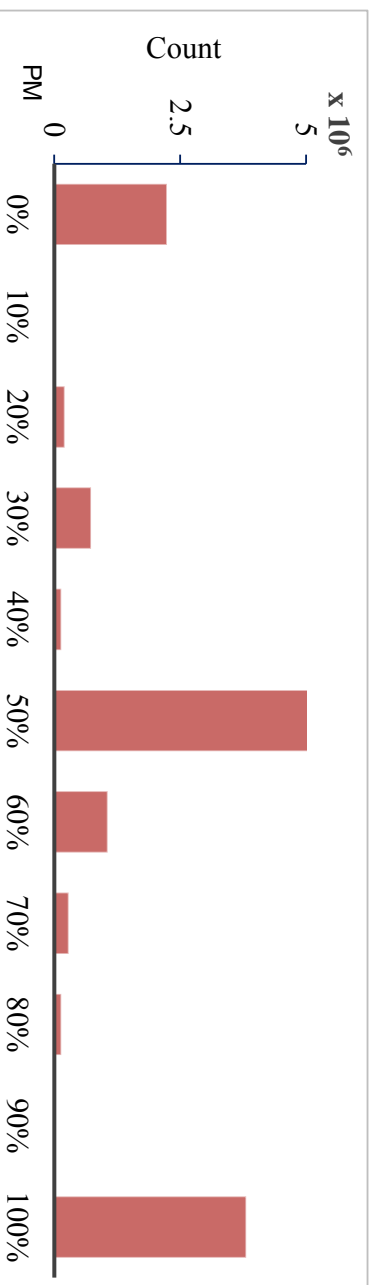


FIGURE 3D. DISTRIBUTIONS OF FUZZY MATCHING PM<sub>2</sub> AMONG FRIENDS (WITH AT LEAST 2 COMMON REVIEWS).

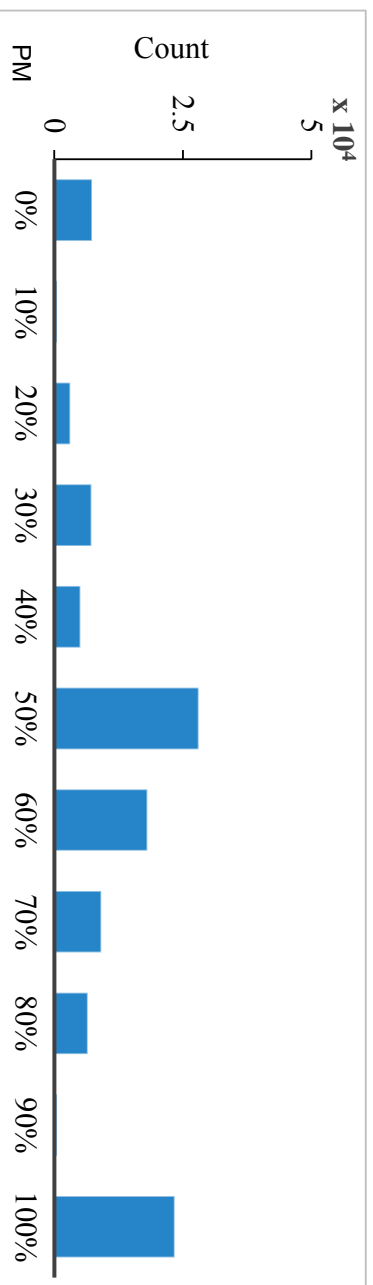


TABLE 1. RESULTS FOR STUDY 1. MEDIAN AND SD IN PARENTHESES.

	Total	Condition	
		60%	90%
(%)			
Actual PM+ displayed	75.34 (88.00, 15.79)	59.39 (60.00, 2.87)	90.33 (90.00, 2.25)
PM+ PPM+	69.91 (70.00, 16.84)	64.23 (65.00, 12.44)	75.24 (80.00, 18.78)
Deviation of PPM+ from PM+	-5.44 (-.50, 18.32)	4.84 (3.00, 11.71)	-15.09 (-10.00, 18.27)
Absolute Deviation	12.75 (9.00, 14.16)	9.29 (8.00, 8.49)	16.00 (10.00, 17.46)
Actual PM- displayed	79.53 (85.00, 18.24)	64.71 (60.00, 14.62)	93.45 (92.00, 6.48)
PPM-	60.95 (60.00, 22.11)	54.68 (59.00, 17.81)	66.85 (70.00, 24.31)
Deviation of PPM- from PM-	-18.58 (-10.00, 25.10)	-10.03 (-6.00, 22.50)	-26.61 (-21.00, 25.07)
PM- Absolute Deviation	22.33 (16.50, 21.78)	16.74 (9.00, 17.91)	27.58 (21.00, 23.96)

TABLE 2. RESULTS FOR STUDY 2. MEDIAN AND SD IN PARENTHESES.

(%)	No info	Condition	
		60%	90%
BR <sub>explicit+</sub>	59.02 (66.00, 25.55)	57.50 (61.50, 26.34)	64.40 (67.50, 18.94)
BR <sub>explicit+</sub> Scaled	64.26 (70.30, 25.83)	62.04 (70, 26.62)	68.09 (70.35, 19.52)
BR <sub>implicit+</sub>	71.36 (80.00, 28.80)	75.32 (90.00, 27.39)	75.69 (80.00, 22.64)
PPM+	66.10 (70.00, 20.19)	62.00 (60.00, 16.29)	81.24 (85.00, 15.40)
PPM-	45.56 (49.00, 19.38)	53.13 (52.50, 23.43)	64.76 (74.50, 27.67)
PM+	-	60.12 (60.00, 13.82)	87.72 (87.50, 5.67)
PM-	-	62.36 (62.50, 33.25)	80.67 (100.00, 29.52)
Bayesian <sub>ppm+</sub>	66.28 (75, 28.72)	67.01 (71.48, 25.42)	82.01 (90, 19.30)

Bayesian <sub>PM+</sub>	-	68.52 (77.78, 25.24)	93.61 (95.53, 6.45)
Actual Forecast+	69.46 (76.00, 25.18)	71.21 (75.50, 22.44)	80.41 (89.50, 21.17)
Actual Forecast+ Scaled	71.06 (80, 25.03)	74.38 (79.07, 21.22)	80.21 (88.87, 21.38)
Weighting Error+	7.99 (10, 17.68)	8.36 (4.29, 21.78)	.36 (0, 17.07)
Forecast Error+	3.55 (2.22, 18.20)	6.80 (4.56, 22.60)	-12.31 (-6.77, 18.59)

TABLE 3. RESULTS FOR STUDY 3. MEDIAN AND SD IN PARENTHESES.

Valence	(%)	No info	Condition		
			60%	90%	
+	Base Rate <sub>implicit+</sub>	80.93 (85.00, 20.33)	75.00 (80.00, 26.18)	79.35 (80.00, 17.37)	
	Base Rate <sub>explicit+</sub>	71.58 (80.00, 23.75)	66.91 (75.00, 26.92)	73.14 (75.00, 18.99)	
	Base Rate <sub>explicit:Scaled+</sub>	68.84 (80.00, 26.30)	62.60 (70.50, 29.75)	63.23 (75.00, 32.53)	
	PPM+	74.09 (77.50, 14.85)	62.10 (60.00, 16.96)	83.72 (90.00, 14.57)	
	PM+	n.a.	58.69 (60.00, 13.95)	87.92 (88.89, 5.84)	
	Explicit Bayesian <sub>PPM+</sub>	77.86 (84.19, 23.41)	74.69 (80.91, 27.01)	88.55 (91.76, 12.52)	
	Explicit Bayesian <sub>PM+</sub>	n.a.	73.25 (83.86, 27.23)	95.75 (97.30, 4.34)	
	Actual Forecast <sub>scaled+</sub>	79.32 (84.31, 20.84)	75.86 (80.57, 22.96)	85.95 (90.00, 13.62)	
	Base Rate <sub>implicit-</sub>	19.07 (15.00, 20.33)	25.00 (20.00, 26.18)	20.65 (20.00, 17.37)	
	Base Rate <sub>explicit-</sub>	26.20 (20.00, 23.17)	30.18 (20.00, 27.03)	25.10 (20.00, 19.49)	
-	Base Rate <sub>explicit:Scaled-</sub>	31.16 (20.00, 26.30)	37.40 (29.50, 29.75)	36.77 (25.00, 32.53)	
	PPM-	48.55 (50.00, 21.75)	54.88 (57.50, 24.35)	61.42 (80.00, 33.04)	
	PM-	n.a.	57.75 (60.00, 33.82)	94.89 (100.00, 15.93)	
	Explicit Bayesian <sub>PPM-</sub>	36.68 (33.67, 26.44)	34.80 (29.64, 29.35)	50.06 (50.84, 27.91)	
	Explicit Bayesian <sub>PM-</sub>	n.a.	35.90 (30.30, 26.64)	66.17 (69.53, 22.55)	
	Actual Forecast <sub>scaled-</sub>	45.03 (50.00, 26.60)	49.06 (50.00, 30.34)	64.36 (74.81, 26.94)	

TABLE 4. RESULTS FOR STUDY 4. MEDIAN AND SD IN PARENTHESES.

Valence	Condition	Condition		
		No info	60%	90%
-	(%)			
	Base Rate <sub>explicit+</sub>	56.82 (60.50, 23.30)	57.31 (60.00, 23.71)	58.48 (61.50, 22.15)
	Base Rate <sub>explicit-Scalcd+</sub>	59.01 (65.00, 23.03)	58.90 (63.87, 23.25)	60.78 (65.00, 21.91)
	PPM+	66.09 (70.00, 17.12)	65.06 (65.00, 14.91)	79.19 (85.00, 17.19)
	PM+	n.a.	59.94 (60.00, 4.25)	90.58 (90.00, 3.01)
	Explicit Bayesian <sub>ppm+</sub> 1	60.32 (68.07, 26.15)	70.50 (76.87, 20.66)	81.23 (89.47, 21.01)
	Explicit Bayesian <sub>ppm+</sub> 1	n.a.	67.90 (77.12, 20.96)	91.95 (95.35, 8.82)
	Explicit Bayesian <sub>ppm+</sub> 2	65.83 (72.73, 24.63)	65.32 (72.44, 25.04)	73.41 (80.19, 25.78)
	Explicit Bayesian <sub>ppm+</sub> 2	n.a.	64.61 (70.32, 23.97)	88.23 (93.49, 14.74)
	Actual Forecast <sub>scaled+</sub> 1	61.39 (60.00, 20.78)	66.31 (68.31, 16.44)	77.68 (81.32, 16.68)
	Actual Forecast <sub>scaled+</sub> 2	65.79 (68.09, 17.80)	70.44 (73.06, 16.91)	73.26 (76.16, 18.23)
	+	Base Rate <sub>explicit-</sub>	38.95 (33.50, 21.83)	39.81 (34.00, 23.06)
Base Rate <sub>explicit-Scalcd-</sub>		40.99 (35.00, 23.03)	41.10 (36.13, 23.25)	39.22 (35.00, 21.91)
PPM-		45.77 (43.00, 21.18)	57.79 (60.00, 20.18)	68.51 (80.00, 25.15)
PM-		n.a.	63.20 (60.00, 13.26)	92.74 (90.00, 6.49)
Explicit Bayesian <sub>ppm-</sub> 1		46.29 (41.48, 23.29)	52.39 (49.60, 27.36)	65.30 (69.23, 24.95)
Explicit Bayesian <sub>ppm-</sub> 1		n.a.	50.36 (48.70, 24.17)	81.03 (84.89, 13.84)
Explicit Bayesian <sub>ppm-</sub> 2		49.34 (49.68, 24.64)	48.60 (41.54, 26.53)	65.91 (68.55, 23.12)
Explicit Bayesian <sub>ppm-</sub> 2		n.a.	47.52 (40.02, 22.18)	79.37 (79.80, 13.26)
Actual Forecast <sub>scaled-</sub> 1		58.36 (57.14, 22.22)	62.31 (60.30, 22.04)	68.44 (70.00, 20.27)
Actual Forecast <sub>scaled-</sub> 2		61.13 (60.00, 20.35)	58.58 (55.28, 17.96)	71.75 (73.73, 18.42)

## REFERENCES

- Ajzen, Icek and Martin Fishbein (1975), "A Bayesian Analysis of Attribution Processes.," *Psychological Bulletin*, 82 (2), 261–77.
- Bar-Hillel, Maya (1980), "The Base-Rate Fallacy in Probability Judgments," *Acta Psychologica*, 44, 211–33.
- Bonaccio, Silvia and Reeshad S. Dalal (2006), "Advice taking and decision-making: An integrative literature review, and implications for the organizational sciences," *Organizational Behavior and Human Decision Processes*, 101 (2), 127–51.
- Bonhard, P. and M. A. Sasse (2006), "'Knowing me, knowing you' - Using profiles and social networking to improve recommender systems," *BT Technology Journal*, 24 (3), 84–98.
- Brewer, Marilyn B. (1979), "In-group bias in the minimal intergroup situation: A cognitive-motivational analysis," *Psychological Bulletin*, 86 (2), 307–24.
- Brown, J O, Amanda J Broderick, and Nick Lee (2007), "Online Communities : Conceptualizing the Online Social Network," *Journal of Interactive Marketing*, 21 (3), 2–20.
- Campbell, Jennifer D. (1986), "Similarity and Uniqueness: The Effects of Attribute Type, Relevance, and Individual Differences in Self-Esteem and Depression," *Journal of Personality and Social Psychology*, 50 (2), 281–94.
- Chen, Yubo, Q I Wang, and Jinhong Xie (2011), "Online Social Interactions : A Natural Experiment on Word of Mouth Versus Observational Learning," *Journal of Marketing Research*, XLVIII, 238–54.

- and Jinhong Xie (2008), “Online Consumer Review : Word-of-Mouth as a New Element of Marketing Communication Mix,” *Management Science*, 54 (3), 477–91.
- Chen, Zoey and Nicholas H. Lurie (2013), “Temporal Contiguity and Negativity Bias in the Impact of Online Word of Mouth,” *Journal of Marketing Research*, L (August), 463–76.
- Chernev, Alexander (2001), “The Impact of Common Features on Consumer Preferences: A Case of Confirmatory Reasoning,” *Journal of Consumer Research*, 27 (4), 475–88.
- Chevalier, Judith A and Dina Mayzlin (2006), “The Effect of Word of Mouth on Sales : Online Book Reviews,” *Journal of Marketing Research*, 43 (3), 345–54.
- Cosmides, Leda and John Tooby (1996), “Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty,” *Cognition*, 58, 1–73.
- Dai, Hengchen, Cindy Chan, and Cassie Mogilner (2015), ““Don’t Tell Me What to Do!’ Shoppers Rely Less on Consumer Reviews for Experiential than Material Purchases.”
- DeBruine, Lisa M (2002), “Facial resemblance enhances trust,” *Proceedings of the Royal Society B: Biological Sciences*, 269, 1307–12.
- Duan, Wenjing, Bin Gu, and Andrew B. Whinston (2008), “The Dynamics of Online Word-of-Mouth and Product Sales - An Empirical Investigation of the Movie Industry,” *Journal of Retailing*, 84 (2), 233–42.

- Eggleston, Casey M, Timothy D Wilson, Minha Lee, and Daniel T Gilbert (2015),  
“Predicting what we will like : Asking a stranger can be as good as asking a friend,”  
*Organizational Behavior and Human Decision Processes*, 128, 1–10.
- Festinger, Leon (1954), “A Theory of Social Comparison Processes,” *Human Relations*,  
7, 117–40.
- Fischer, Peter, Stefan Schulz-Hardt, and Dieter Frey (2008), “Selective exposure and  
information quantity: how different information quantities moderate decision  
makers’ preference for consistent and inconsistent information.,” *Journal of  
personality and social psychology*, 94 (2), 231–44.
- Freedman, J L (1965), “Preference for Dissonant Information,” *Journal of personality  
and social psychology*, 56, 287–89.
- Gershoff, Andrew D, Ashesh Mukherjee, and Anirban Mukhopadhyay (2003),  
“Consumer Acceptance of Online Agent Advice: Extremity and Positivity Effects,”  
*Journal of Consumer Psychology*, 13, 161–70.
- Gershoff, Andrew D., Ashesh Mukherjee, and Anirban Mukhopadhyay (2008), “What ’s  
Not to Like ? Preference Asymmetry in the False Consensus Effect,” *Journal of  
Consumer Research*, 35 (1), 119–25.
- 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.
- Gigerenzer, Gerd and Ulrich Hoffrage (1995), “How to Improve Bayesian Reasoning  
Without Instruction : Frequency Formats,” *Psychological Review*, 102 (4), 684–704.

- Gilbert, D T, E C Pinel, T D Wilson, S J Blumberg, and T P Wheatley (1998), "Immune neglect: a source of durability bias in affective forecasting.," *Journal of personality and social psychology*, 75 (3), 617–38.
- Gilbert, Daniel T and Jane E J Ebert (2002), "Decisions and revisions: the affective forecasting of changeable outcomes.," *Journal of personality and social psychology*, 82 (4), 503–14.
- Gino, Francesca and Maurice E Schweitzer (2008), "Blinded by anger or feeling the love: how emotions influence advice taking.," *The Journal of applied psychology*, 93 (5), 1165–73.
- Godes, D. and D. Mayzlin (2009), "Firm-Created Word-of-Mouth Communication: Evidence from a Field Test," *Marketing Science*, 28 (4), 721–39.
- Goodie, Adam S and Edmund Fantino (1999), "What Does and Does Not Alleviate Base-Rate Neglect Under Direct Experience," *Journal of Behavioral Decision Making*, 12, 307–35.
- He, Stephen X. and Samuel D. Bond (2013), "Word-of-mouth and the forecasting of consumption enjoyment," *Journal of Consumer Psychology*, 23 (4), 464–82.
- and ——— (2015), "Why Is the Crowd Divided? Attribution for Dispersion in Online Word of Mouth," *Journal of Consumer Research*, 41 (6), 1509–27.
- Hoffrage, Ulrich and Gerd Gigerenzer (1998), "Using natural frequencies to improve diagnostic inferences," *Academic Medicine*, 73 (5), 538–40.



- , Stefan Krauss, Laura Martignon, and Gerd Gigerenzer (2015), “Natural frequencies improve Bayesian reasoning in simple and complex inference tasks,” *Frontiers in Psychology*, 6 (October), 1–14.
- Kahneman, D and A Tversky (1973), “On The Psychology of Prediction,” *Psychological Review*, 80 (4), 237–51.
- Kanouse, David E (1984), “Explaining Negativity Biases in Evaluation and Choice Behavior: Theory and Research,” *Advances in Consumer Research*, 11, 703–8.
- Kermer, Deborah a, Erin Driver-Linn, Timothy D Wilson, and Daniel T Gilbert (2006), “Loss aversion is an affective forecasting error.,” *Psychological science*, 17 (8), 649–53.
- Kleck, R E and J Wheaton (1967), “Dogmatism and Responses to Opinion-Consistent and Opinion-Inconsistent Information.,” *Journal of personality and social psychology*, 5 (2), 249–52.
- Koehler, Jonathan J (1996), “The Base Rate Fallacy Reconsidered : Descriptive , Normative , and Methodological Challenges,” *Behavioral and Brain Sciences*, 19, 1–53.
- Linden, Greg, Brent Smith, and Jeremy York (2003), “Amazon.com recommendations: Item-to-item collaborative filtering,” *IEEE Internet Computing*, 7 (1), 76–80.
- Liu, Yong (2006), “Word of Mouth for Movies: Its Dynamics and Impact on Box Office Revenue,” *Journal of Marketing*, 70 (3), 74–89.
- Loewenstein, George, Ted O’Donoghue, and Matthew Rabin (2003), “Projection Bias in Predicting Future Utility,” *The Quarterly Journal of Economics*, 118, 1209–48.

- Lord, Charles G, Lee Ross, and Mark R Lepper (1979), “Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence.,” *Journal of Personality and Social Psychology*, 37 (11), 2098–2109.
- Luca, Michael and G Zervas (2015), “Fake It Till You Make It: Reputation, Competition, and Yelp Review Fraud,” *Harvard Business School NOM Unit Working Paper No. 14-006*, 1–35.
- Lynn, Michael and Judy Harris (1997), “Individual Differences in the Pursuit of Self-Uniqueness Through Consumption,” *Journal of Applied Social Psychology*, 27 (21), 1861–83.
- Mahajan, Neha and Karen Wynn (2012), “Origins of ‘us’ versus ‘them’: prelinguistic infants prefer similar others,” *Cognition*, 124 (2), 227–33.
- Marks, Gary and Norman Miller (1987), “Ten Years of Research on the False-Consensus Effect : An Empirical and Theoretical Review,” *Psychological Bulletin*, 102 (1), 72–90.
- Mcpherson, Miller, Lynn Smith-lovin, and James M Cook (2001), “Birds of a Feather : Homophily in Social Networks,” *Annual Review of Sociology*, 27, 415–44.
- Medin, D L and S M Edelson (1988), “Problem structure and the use of base-rate information from experience.,” *Journal of experimental psychology. General*, 117 (1), 68–85.
- Mellers, Barbara A. and A. Peter McGraw (2001), “Anticipated Emotions as Guides to Choice,” *Current Directions in Psychological Science*, 10 (6), 210–14.

- Meng, Rachel, Stephanie Y. Chen, and Daniel M. Bartels (2016), "Reasoning About Preference Diversity When Taking and Seeking Advice."
- Mudambi, Susan M and David Schuff (2010), "What Makes a Helpful Online Review? A Study of Customer Reviews on Amazon.com," *MIS Quarterly*, 34 (1), 185–200.
- Mussweiler, T. (2001), "Focus of Comparison as a Determinant of Assimilation Versus Contrast in Social Comparison," *Personality and Social Psychology Bulletin*, 27 (1), 38–47.
- Naylor, Rebecca Walker, Cait Poyner Lamberton, and David a Norton (2011), "Seeing Ourselves in Others: Reviewer Ambiguity, Egocentric Anchoring, and Persuasion," *Journal of Marketing Research*, 48 (3), 617–31.
- Newby-clark, Ian R, Ian Mcgregor, and Mark P Zanna (2002), "Thinking and Caring About Cognitive Inconsistency : When and for Whom Does Attitudinal Ambivalence Feel Uncomfortable ?," *Journal of Personality and Social Psychology*, 82 (2), 157–66.
- Olshavsky, Richard W, John A Miller, Richard W Olshavsky, and John A Miller (1972), "Consumer Expectations , Product Performance , and Perceived Product Quality," *Journal of Marketing Research*, 9 (1), 19–21.
- Osgood, Charles E and Percy H Tannenbaum (1955), "The Principle of Congruity in The Prediction of Attitude Change," *Psychological Review*, 62 (1), 42–55.
- Patrick, Vanessa M., Deborah J. MacInnis, and C. Whan Park (2007), "Not as Happy as I Thought I'd Be? Affective Misforecasting and Product Evaluations," *Journal of Consumer Research*, 33 (4), 479–89.

- Price, Linda L., Lawrence F. Feick, and Robin a. Higue (1989), "Preference heterogeneity and coorientation as determinants of perceived informational influence," *Journal of Business Research*, 19 (3), 227–42.
- 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, 219–301.
- Schkade, D. a. and D. Kahneman (1998), "Does Living in California Make People Happy? A Focusing Illusion in Judgments of Life Satisfaction," *Psychological Science*, 9 (5), 340–46.
- Schulz-Hardt, Stefan, Dieter Frey, Carsten Lüthgens, and Serge Moscovici (2000), "Biased information search in group decision making," *Journal of Personality and Social Psychology*, 78 (4), 655–69.
- Scott, Carol A and Richard F Yalch (1980), "Consumer Response to Initial Product Trial: A Bayesian Analysis," *Journal of Consumer Research*, 7 (1), 32–41.
- Simons, Herbert W, Nancy N Berkowitz, and R John Moyer (1970), "Similarity, Credibility, and Attitude Change: A Review and a Theory," *Psychological Bulletin*, 73 (1).
- Skowronski, JJ and DE Carlston (1987), "Social judgment and social memory: The role of cue diagnosticity in negativity, positivity, and extremity biases.," *Journal of Personality and Social Psychology*, 52 (4), 689–99.

- Snizek, Janet A, Gunnar E Schrah, and Reeshad S Dalal (2004), "Improving judgement with prepaid expert advice," *Journal of Behavioral Decision Making*, 17 (June), 173–90.
- Snizek, Janet a. and Lyn M. Van Swol (2001), "Trust, Confidence, and Expertise in a Judge-Advisor System.," *Organizational behavior and human decision processes*, 84 (2), 288–307.
- Snyder, C R and H L Fromkin (1977), "Abnormality as a Positive Characteristic : The Development and Validation of a Scale Measuring Need for Uniqueness," *Journal of Abnormal Psychology*, 86 (5), 518–27.
- Sood, Ashish, Mayukh Dass, Wolfgang Jank, and Yue Tian (2016), "Power of Customer Voice: Shape Analysis of Online Product Reviews."
- Srull, Thomas K. and Lisa Gaelick (1983), "General Principles and Individual Differences in the Self as a Habitual Reference Point: An Examination of Self-Other Judgments of Similarity," *Social Cognition*.
- Su, X. and T. M. Khoshgoftaar (2009), "A Survey of Collaborative Filtering Techniques," *Advances in Artificial Intelligence*, 2009, 1–19.
- Suls, J, R Martin, and L Wheeler (2002), "Social comparison: Why, with whom, and with what effect?," *Current Directions in Psychological Science*, 11 (5), 159–63.
- Suls, Jerry, Rene Martin, and Ladd Wheeler (2000), "Three Kinds of Opinion Comparison: The Triadic Model," *Personality and Social Psychology Review*, 4 (3), 219–37.

- and C. K. Wan (1987), “In Search of the False-Uniqueness Phenomenon: Fear and Estimates of Social Consensus,” *Journal of Personality and Social Psychology*, 52 (1), 211–17.
- Tian, Kelly Tepper, William O. Bearden, and Gary L. Hunter (2001), “Consumers’ need for uniqueness: scale development and validation,” *Journal of Consumer Research*, 28 (6), 50–66.
- Trope, Y. and M. Bassok (1982), “Confirmatory and diagnosing strategies in social information gathering,” *Journal of Personality and Social Psychology*, 43 (1), 22–34.
- Trope, Yaacov (1974), “Inferential Processes in the Forced Compliance Situation: A Bayesian Analysis,” *Journal of Experimental Social Psychology*, 10, 1–16.
- and Eugene Burnstein (1975), “Processing the Information Contained in Another’s Behavior,” *Journal of Experimental Social Psychology*, 11, 439–58.
- Trusov, Michael, Randolph E Bucklin, and Koen Pauwels (2009), “Effects of Word-of-Mouth Versus Traditional Marketing: Findings from an Internet Social Networking Site,” *Journal of Marketing*.
- Tversky, A and D Kahneman (1974), “Judgment under Uncertainty: Heuristics and Biases.,” *Science (New York, N.Y.)*, 185 (4157), 1124–31.
- Twyman, M, N Harvey, and C Harries (2008), “Trust in motives, trust in competence: Separate factors determining the effectiveness of risk communication,” *Judgment and Decision Making*, 3 (1), 111–20.

- Walsh, Emma and Peter Ayton (2009), “My imagination versus your feelings: can personal affective forecasts be improved by knowing other peoples’ emotions?,” *Journal of experimental psychology. Applied*, 15 (4), 351–60.
- Wilson, T D, T Wheatley, J M Meyers, D T Gilbert, and D Axsom (2000), “Focalism: A Source of Durability Bias in Affective Forecasting,” *Journal of personality and social psychology*, 78 (5), 821–36.
- Wilson, Timothy D and Daniel T Gilbert (2005), “Affective Forecasting : Knowing What to Want,” *Current Directions in Psychological Science*, 14 (3), 131–34.
- Wu, Philip Fei (2013), “In Search of Negativity Bias: An Empirical Study of Perceived Helpfulness of Online Reviews,” *Psychology & Marketing*, 30 (11), 971–84.
- Yaniv, Ilan (2004), “The Benefit of Additional Opinions,” *Current Directions in Psychological Science*, 13 (2), 75–78.
- , Shoham Choshen-Hillel, and Maxim Milyavsky (2011), “Receiving advice on matters of taste: Similarity, majority influence, and taste discrimination,” *Organizational Behavior and Human Decision Processes*, 115 (1), 111–20.
- and Maxim Milyavsky (2007), “Using advice from multiple sources to revise and improve judgments,” *Organizational Behavior and Human Decision Processes*, 103, 104–20.
- Yates, J. Frank, Paul C. Price, Ju-Whei Lee, and James Ramirez (1996), “Good Probabilistic Forecasters : The ‘Consumer’ s’ Perspective,” *International Journal of Forecasting*, 12, 41–56.

- Yoon, Yeosun, Gulen Sarial-Abi, and Zeynep Gurhan-Canli (2012), “Effect of Regulatory Focus on Selective,” *Journal of Consumer Research*, 39 (June), 93–110.
- Zanna, Mark P and Mark R Lepper (1973), “Attentional Mechanisms in Children ’ S Devaluation of a Forbidden Activity in a Forced-Compliance Situation,” *Journal of Personality and Social Psychology*, 28 (3), 355–59.
- Zhu, Feng and Xiaoquan (Michael) Zhang (2010), “Impact of Online Consumer Reviews on Sales: The Moderating Role of Product and Consumer Characteristics,” *Journal of Marketing*, 74 (2), 133–48.