Valuing Dissimilarity: the Role of Diversity in Preference Predictions

Rachel Meng, Columbia University, USA
Stephanie Y. Chen, University of Chicago, USA
Daniel M. Bartels, University of Chicago, USA

We examine how people integrate opinions from similar and dissimilar others to predict matters of taste. People prefer to sample and use information from diverse (vs. similar) advisors when making more verifiable judgments about an unfamiliar item and when they perceive a product category to represent matters of objective quality.

[to cite]:

[url]:
http://www.acrwebsite.org/volumes/1020335/volumes/v43/NA-43

[copyright notice]:
This work is copyrighted by The Association for Consumer Research. For permission to copy or use this work in whole or in part, please contact the Copyright Clearance Center at http://www.copyright.com/.
Advances in Consumer Research
Volume 43, ©2015

Reasoning About Advice: Inferring and Integrating the Preferences of Others
Chair: Rachel Meng (Columbia University)

Paper #1: Mistaking Dissimilar for Dislike: Why We Underestimate the Diversity of Others’ Preferences
Kate Barasz, Harvard University, USA
Tami Kim, Harvard University, USA
Leslie John, Harvard University, USA

Paper #2: Valuing Dissimilarity: The Role of Diversity in Preference Predictions
Rachel Meng, Columbia University, USA
Stephanie Y. Chen, University of Chicago, USA
Daniel M. Bartels, University of Chicago, USA

Paper #3: Are Advice Takers Bayesian? Preference Similarity Effects on Advice Seeking and Taking
Hang Shen, University of California, Riverside, USA
Ye Li, University of California, Riverside, USA

Paper #4: Recommenders vs. Recommender Systems
Mike Yeomans, Harvard University, USA
Anuj Shah, University of Chicago, USA
Sendhil Mullainathan, Harvard University, USA
Jon Kleinberg, Cornell University, USA

SESSION OVERVIEW
The opinions of others often shape our beliefs about the world; these beliefs can in turn influence our inferences, predictions, and choices. In the information age we inhabit nowadays, the vast quantity of advice accessible to us—be it through online recommendation systems, social media, professionals and experts, or family and friends—renders it impractical to make decisions alone. Yet, research on advice and the wisdom of crowds (Larrick et al. 2012; Yaniv 2004) finds that consumers fail to fully exploit this wisdom. To address this, the proposed session discusses how people reason about advice in matters of taste: inferring others’ preferences (Papers 1-2), seeking and incorporating recommendations to make predictions (Papers 2-3), and leveraging human and nonhuman sources of advice to ultimately improve prediction accuracy (Paper 3-4). In particular, the featured papers illuminate a range of faulty inferences that impede this goal.

Barasz, Kim, and John (Paper 1) begin by articulating a systematic prediction error when forming inferences about others’ tastes. This “preference homogeneity bias,” in which consumers believe others to possess less diverse preferences than themselves, may consequently affect how they seek and take advice from others. Next, Meng, Chen, and Bartels (Paper 2) examine how and when information from diverse others is used in inductive reasoning. Compared to predicting their own tastes, people were willing to sample advice from both similar and dissimilar others when they made more verifiable judgments and when they perceived a product category to represent matters of objective quality. Shen and Li (Paper 3) expand on the role of similarity by exploring how people value preference matching with online surrogates (users who have experienced a product) to guide preference predictions. They found a tendency to undervalue preference matching but overweight surrogate advice, indicating a departure from Bayesian standards. Finally, Yeomans, Shah, Mullainathan, and Kleinberg (Paper 4) highlight a persistent aversion to algorithmic advice. Despite the superiority of collaborative filtering in accurately predicting tastes, humans remain skeptical about such recommendations.

In concert, these papers call attention to ways in which individuals commit prediction errors involving matters of taste. People falsely assume their own preferences to be more diverse than those of others (Barasz, Kim, and John) and are reluctant to seek diversity when predicting their future utility (Meng, Chen, and Bartels). They neglect base rates when taking surrogate advice (Shen and Li) and remain unresponsive to innovative sources of advice which can improve decisions and reduce search costs (Yeomans et al.). However, people do take into account others’ opinions when nudged (Barasz, Kim, and John; Meng, Chen, and Bartels).

Listening to others is something we do habitually, and a closer examination of when and why we fail at doing so is crucial if we are to recommend effective remedies to these errors. Borrowing insights from multiple perspectives that span a variety of consumption domains, the substantive issues raised in this session may be especially useful for those interested in affective forecasting, inductive reasoning, advice and the wisdom of crowds, preferences, social influence, and word-of-mouth.

Mistaking Dissimilar for Dislike: Why We Underestimate the Diversity of Others’ Preferences

EXTENDED ABSTRACT
Suppose you are asked to predict someone’s preferences for certain items in a given category—say, types of vacation destinations or movie genres. Would the person like traveling to a bustling city? Would he enjoy watching an action-packed thriller? Lacking any specific information about that person’s tastes, you might evaluate the items on their own merits, and for any popular or likeable options, sensibly predict that the other person would enjoy them.

But what if you had access to this other person’s tastes for other items in the same category? Perhaps you discover the person recently vacationed at a lake house, or just attended the premiere of a documentary film. How might this change your predictions? Knowing the other person had previously chosen options seemingly unlike city vacations or thriller films, you might conclude he would not enjoy either. As this paper explores, upon finding out about others’ preferences for one option, people have a recurring tendency to predict that others will dislike dissimilar options within the same category.

Is this inference correct? After observing another person’s choice, we not only make assumptions about how much she likes the chosen option (Miller and Nelson 2002); we also make broader inferences about how much she likes—or dislikes—unchosen options. After learning of someone’s choice—or their “background preference”—we show that people erroneously expect that others will dislike dissimilar ones. However, people hold this expectation despite recognizing that they, themselves, simultaneously like dissimilar options. For example, people readily indicate enjoying dissimilar vacation destinations (e.g., lake and city) and dissimilar movies (e.g., documentaries and thrillers), but predict that—for others—a preference for one precludes enjoyment of the other. Simply put, when predicting others’ preferences, people mistake dissimilarity for dislike. In five experiments, we document this prediction error and show that it is driven by a false belief that others have a narrower, more homogeneous range of preferences than ourselves.

Study 1 demonstrates the basic effect: People mistakenly believe that others do not like items that are dissimilar from one another. Participants read about a consumer choosing between three
vacation destinations: Lake, Mountain, or City—the first two were similar, whereas the latter was dissimilar. Participants learned about the consumer’s preferences for a reference choice (lake)—either that she chose Lake, ruled out Lake, or were given no information—and then estimated how much she liked each of the three options. Participants estimated positively correlated liking levels for similar options ($r=.56, p<0.001$), but negatively correlated levels for dissimilar options ($r=-.42, p<0.001$), as though these dissimilar preferences were mutually exclusive. In contrast, when reporting their own preferences, people’s own preferences for dissimilar options were far less polarized ($r=-.10, p=0.054$). Thus, Study 1 offers preliminary evidence that predictions and base rates are not aligned for dissimilar items.

Study 2 demonstrates the prediction error using a measure of dichotomous choice, shows it is specific to dissimilar items, and varies product ratings to gauge (actual/predicted) tradeoffs between quality and similarity. In a 2 (perspective: actors vs. observers) × 2 (options: similar vs. dissimilar) design, participants either chose a movie genre for themselves (actors) or learned of someone else’s choice (observers), as between two similar options (thriller/action adventure) or two dissimilar ones (thriller/documentary). Then participants made an “out-of-stock” tradeoff choice: a 3-star version of the chosen genre (e.g., 3-star thriller), or a 5-star version of the alternative (e.g., 5-star action adventure or 5-star documentary) genre. We predicted that, for themselves, most actors would rather have the higher-rated movie, regardless of genre; however, when predicting someone else’s choice, we believed observers would correctly anticipate a tradeoff between similar items, but—mistaking dissimilar for dislike—would not predict a tradeoff between dissimilar ones. Indeed, when options were similar, most actors (74%) and observers (69%) chose the similar, higher-rated movie. However, when options were dissimilar, a significant gap emerged: most actors chose the dissimilar, higher-rated movie (64%), a choice predicted by few observers (18%; $p<.001$).

Study 3 elucidates the process underlying the effect: mistaking dissimilar for dislike is fully mediated by a (false) belief that others’ preferences are homogeneous. We replicated the same design as Study 2’s dissimilar condition, having people make a tradeoff (for themselves or someone else) between a 3-star similar movie and a 5-star dissimilar one. In addition, we measured actual/perceived range of preferences. Drawing on out-group homogeneity research (Judd, Ryan, & Park 1991), we showed participants 14 movie genres and asked them to select all the ones they (actors) or someone else (observers) liked. The results show that people infer a narrower, more homogeneous range of preferences for others: while actors reported liking 7 movie genres, observers estimated that someone else only liked 4. This fully mediated the choice of dissimilar movie in the out-of-stock scenario.

Study 4 shows that the effect persists even when predictors are financially incentivized for accuracy. Lab participants were partnered, assigned to be actors or observers. Actors chose between two dissimilar movies; observers learned of this choice. Both partners then encountered Study 2’s trade-off choice. Observers were promised a bonus for correct prediction. While most actors (69%) chose the higher-rated but dissimilar option for themselves, again few observers (39%; $p<0.001$) predicted that choice.

Finally, Study 5 found that the preference homogeneity bias is tempered by encouraging people to consider that others’ preferences are varied (Critcher and Dunning 2013). Here, participants learned that either 1 person or 100 people had chosen a movie (thriller/documentary), then predicted a choice between a 3-star similar movie and a 5-star dissimilar one. Participants in the single person condition selected the choice they believed the other person made, while those in the sample population condition estimated how many of the 100 people chose 3-star similar and how many chose 5-star dissimilar. Replicating previous results, only 27% of participants predicted that a single person would choose the higher-quality, dissimilar movie; however, for the sample population, observers predicted that 53% would choose the higher-quality, dissimilar movie, suggesting that the error can be mitigated when thinking more globally.

Valuing Dissimilarity: The Role of Diversity in Preference Predictions

EXTENDED ABSTRACT

People frequently incorporate the opinions of others to make predictions about the world, including their preferences for novel experiences. Consider a scenario where a person is deciding whether to see a new movie she knows little about. To predict how much she will enjoy it, she can solicit others’ opinions. But whose advice does she value more—that of people with a wide variety of movie tastes, or that of only people who share her tastes? The current research asks how we integrate the opinions of those who are similar and dissimilar to us to inform predictions.

Similar others tend to be more influential on our judgments than dissimilar others (Festinger 1954; Heider 1958; Suls, Martin, and Wheeler 2002). Not only do people treat similar others as reliable and attractive sources of information, but they tend to discount advice from those less like them (Twyman, Harvey, and Harries 2008). However, there is reason to believe that people may value dissimilar others. The diversity principle from category-based induction states that evidence from diverse sources support stronger arguments (Heit 2000). These diverse samples create a stronger basis for generalization because they better cover the category of interest (Osherson et al. 1990). People also prefer to seek diverse, rather than similar, pieces of evidence when judging the validity of generalizations (López 1995; Rhodes, Brickman, and Gelman 2008). For example, when assessing whether an unfamiliar property (has sesamoid bones) holds for all mammals, participants would rather test whether it holds for lions and goats than for lions and leopards.

Further, the social influence literature has shown that people are more swayed by dissimilar others when more verifiable beliefs rather than values are involved (Goethals and Nelson 1973). The studies below explore whether and when people care about evidential diversity when predicting beliefs about preferences (“Will I like this movie?”) as they do when predicting beliefs about facts (“Does a sea cucumber have condyloid canals?”).

In Study 1, 156 MTurk respondents viewed positive evaluations of an unfamiliar movie from a pair of movie-goers, one who was similar to them and another who was dissimilar. When asked to describe how they would use this information to predict their enjoyment for the movie, 61% indicated that the dissimilar movie-goer’s rating strengthened the likelihood that they would like the movie. Only 21% reported attending solely to the similar person. Hence, congruent opinions expressed by a diverse set of individuals may favorably influence predictions.

Study 2 examined whether these explicit self-reports matched how people seek advice to make preference predictions. Participants ($N=201$) completed an evidence selection task where they solicited opinions from a panel of 12 “regular movie-goers” (reviewers) ranked by their similarity to the participant’s movie preferences. We assigned participants to one of four conditions: People were asked to choose three reviewers whose opinions they would most like to get in order to predict (a) how much they would like a target mystery movie, (b) how much the average person would like it, (c) how criti-
ally acclaimed the movie would be, or (d) how successful at the box office it would be. We suspect diversity effects to emerge in the last three conditions, which constitute more verifiable judgments about the target movie.

For each individual, we calculated two indices based on her chosen distribution of reviewers: a “dissimilarity score,” capturing willingness to sample reviewers with less overlap to the participant, and a “diversity score,” capturing willingness to sample reviewers with broader overlap. Participants solicited both more dissimilar and diverse opinions when predicting more verifiable features of the movie than when predicting their own preferences. Whereas 44% of those in the self condition selected an advisor with a rank greater than 6 (higher ranks mean greater dissimilarity), 67%, 68%, and 55% did so in the average person, success, and critical acclaim conditions.

Study 3 (N=161) used an identical procedure, with one exception: Rather than elicit predictions about verifiable features, we varied the product category. Spiller and Belogolova (2015) argue that consumers hold discrepant beliefs about whether a set of products is “vertically differentiated” (differing on quality and intrinsic value, thus more verifiable) or “horizontally differentiated” (differing on taste and idiosyncratic value). A pretest of 35 product categories revealed significant inter-category heterogeneity in perceived differentiation. For example, respondents agreed that paintings, restaurants, and tattoos belonged to the realm of taste, while public transportation, blenders, and digital cameras concerned objective quality. We selected restaurants and digital cameras to represent two opposing ends of the taste-quality spectrum for Study 3. Results found that people sampled more dissimilar and diverse advisors for digital cameras relative to restaurants.

Studies 2-3 suggest that individuals are more likely to seek diversity when forming more verifiable judgments or when they perceive a category as vertically differentiated. Study 4 examined how they then use these opinions to update predictions. We assigned participants (N=398) to one of eight conditions in a 4 (prediction: self, average person, acclaim, success) x 2 (advisor pair: similar, diverse) between-subjects design. Participants saw ratings from a pair of advisors sequentially, making one prediction after seeing each opinion. Half saw two similar reviewers, while the other half saw one similar followed by one dissimilar reviewer. In both conditions, the two advisors rated the movie highly. While participants contrasted their own predictions away from the dissimilar advisor, no contrast effects emerged for predictions pertaining to the average person, critical acclaim, or box office success. Thus, preference diversity appears to a negative cue only for judgments of personal preference.

Taken together, these results have implications for what information people might seek in different contexts. For example, if we were tasked with judging the quality of an ACR paper, we may prefer to poll conference attendees with both similar and dissimilar interests to our own. By contrast, if we wanted to determine which paper we should read for personal pleasure, we may only bother colleagues who share our taste in papers. Identifying precisely which factors affect how broadly we sample advice is an important topic of future research.

Are Advice Takers Bayesian? Preference Similarity Effects on Advice Seeking and Taking

EXTENDED ABSTRACT

Consumers today have access to far more information when making purchasing decisions than ever, thanks mainly to the proliferation of consumer-oriented websites and apps. When consumers consider purchasing new products, they may try to forecast how likely they will like it. To better inform this “probabilistic affective forecast”, they may generally use two broad methods: 1) simulation of what it would be like to use the product based on product descriptions and pictures, and 2) surrogation—seeking word-of-mouth (WOM) or advice from surrogates (i.e., people who have direct experience with the product), for example, by reading their reviews on websites including Yelp, Amazon, and IMDb (Gilbert et al. 2009).

To date, little research has explored the more recent trend of online reviews becoming more social. For example, Yelp allows users to “friend” other users. On the other hand, social media websites such as Facebook and Twitter now provide users with product suggestions based on their friends’ ratings. Consumers now are able to take advice from someone who they at least partially know the preferences of, and thus have more information than before. However, previous research has yet to explore whether this additional information helps consumers make better decisions and rely more on WOM (He and Bond 2013).

The present research examines consumers’ affective forecasts when taking advice (WOM) from a surrogate whose preferences are at least partially known, and what determines whether consumers are willing to seek advice from these surrogates. We explore whether taking advice from online friends helps overcome systematic errors people often make in their affective forecasts of future experiences, findings that are well documented in previous research (Kemrer et al 2009; Gilbert and Wilson 2000; Patrick, MacInnis, and Park 2007). Specifically, we directly manipulated the degree of preference matching (PM) with the surrogate to see how participants incorporate different PM levels into their affective forecasts and how PM affects advice seeking.

To assess the normative degree to which consumers should take advice from surrogates in the presence of PM information, we introduce a Bayesian framework that allows us to evaluate ex-ante prediction accuracy. There are two variables to consider when consumers incorporate a surrogate’s advice to predict their enjoyment of a future experience: 1) their general liking of similar experiences (i.e., the base rate), and 2) the relevance of the surrogate advice, which depends on PM (i.e., the diagnosticity of the advice). Therefore, it seems reasonable to draw an analogy between the advice-taking process and a Bayesian updating paradigm.

Two studies presented here directly manipulated PM with an unknown surrogate and measured the prediction accuracy using Bayesian prediction as the standard. We thus extend the research on affective forecasting, advice taking, and WOM to better understand the advice-taking process in the social media era.

Study 1 tested how perceived PM (PPM) affects consumers taking and further seeking advice from the surrogate. We mimic a complete decision process of making a probabilistic affective forecast (“how likely I will like it?”). Participants directly experienced PM by seeing “actual” ratings from a previous participant after rating each of 10 pieces of artwork, the trials that enable participants to recall their base rate preference and PM history with the surrogate. Then they saw the surrogate advice and made the prediction on the 11th piece of artwork. After seeing the 11th piece of artwork, they were asked to make a choice between surrogation information (previous participant’s rating) and reading a short description of the 12th piece of artwork. We used a 5 (PM: 50% vs. 70% vs. 90% vs. 100% vs. control) X 2 (advice valence: 5 star vs. 1 star) between-subjects design.

We found that PPM is lower than the given PM, except for the 50% condition. This suggests that when consumers consider PM, the default is to assume low matching. Our study measured prediction
accuracy using Bayesian implicit PM (IPM), which is derived back from the actual prediction, representing the PM that a Bayesian predictor would have to reach this prediction, given the base rate. Comparing PPM with IPM reveals overuse of updated information (surrogate rating), as IPM that was actually used for predicting is higher than PPM. This is consistent with our expectation that although consumers perceive lower PM with a surrogate, they make an opposite decision that overweights on surrogate advice.

After experiencing the target, more participants chose surrogation over descriptive information when the surrogate rating on the 11th painting agreed with their actual rating than when contradicted. Therefore, “useful” surrogate advice is particularly important for continuing to go to that surrogate for further advice. This contradicts the finding in previous research (e.g. Gilbert et. al. 2009) that people always prefer descriptive information over surrogate advice.

Study 2 switched domains to Dilbert comics, used a different range of PM (60%, 90%, and control) and changed the timing of choosing advice type. We enforced the choice before seeing the advice and making the prediction.

Participants reported lower-than-given PPM only in 90% condition. The PPM in 60% condition (57.84%) is similar to that in the control condition (57.04%), implying that people’s default of PPM is possibly around 60%. We found no difference in choice of advice type between the control condition and 60% condition, but significantly higher preference for surrogate rating in the 90% condition. A logit regression analysis shows that the PPM is the major predictor of the choice on advice type. Similar to Study 1, participants who chose surrogate rating overweighted surrogate’s advice across all three conditions based on the comparisons between IPMs and PPMs.

In summary, we explored how PM affects consumers using and seeking surrogate advice. To corporations who deal with massive online WOM, this research rationalizes their effort of socializing online reviews but also suggests that it may backfire. Providing PM information could benefit higher acceptance of online reviews, but may significantly reduce the intention of seeking advice once the advice quality is proved to be poor. Our proposition on Bayesian prediction could also help consumers to make better decisions using online reviews.

**RECOMMENDERS VS. RECOMMENDER SYSTEMS**

**EXTENDED ABSTRACT**

Recommendations can have a tremendous impact on welfare by guiding consumers towards better choices. Most often, they have been generated by other humans (experts, peers, loved ones, etc.) but recently, recommendations based on collaborative filtering algorithms have been introduced to many consumption settings (YouTube, Netflix, Amazon, etc.). These systems flourish because the costs of generating predictions are small. But costs aside, how beneficial are they for consumers, compared to the alternative? Can the algorithms make better predictions about preferences than other people? And which type of recommendation would people prefer to receive?

We conducted a series of experiments that compared human and machine recommenders, both on objective accuracy and on subjective consumer taste. We selected a domain where humans ought to have a natural advantage: Predicting which jokes other people will find funny. Strikingly, even in this domain, we find that computer algorithms are better recommenders than people, regardless of whether these people are making recommendations for complete strangers or close others. And yet, people would prefer to receive recommendations from other humans.

In Study 1 we tested whether human recommenders or computer algorithms were better at predicting people’s preferences. Our participants (N=197) predicted how much users of Jester (a website where people rate jokes) would like different jokes. We also developed a collaborative filtering algorithm to make these predictions. Each participant saw 4 jokes that a user had read and rated (e.g. this user rated joke A a 5.2, joke B a 1.4, and so on). Participants then read two new jokes and predicted which joke the user liked more, repeated for five users. On the machine recommender side, the algorithm took as inputs the user ratings. The algorithm then scanned a database of over 4000 Jester users and identified users who had made similar ratings for the four jokes. It then applied a kernel weighting function to develop a prediction of how much the current user would like a joke. Averaging across all five predictions the human subjects guessed correctly less often (M=56%, SE=1.6%) than the algorithm (M=63%, SE=1.5%).

Note that humans have one kind of advantage: They comprehend the subject and punchline of the joke, and they have direct insight into why it is funny. On the other hand, collaborative filtering doesn’t actually understand anything about the content of humor. Instead, they only scan for regularities in the pattern of ratings for those particular jokes. This is different from an actuarial model (Dawes 1979), which tries to reconstruct the human judgment process, using the same inputs. However, there are other sources of information (missing in Study 1) that human recommenders usually rely on in the real world—recommenders often know their targets personally, and know the context in which a joke will be heard. Would this make a difference?

In Study 2, we recruited pairs of people (N=122) who knew each other well (significant others, family members, etc.). Both people first rated 12 jokes for themselves. Then they each turned to making predictions about the other person. They saw a random sample of four jokes, along with their partner’s ratings for those jokes. Then they predicted their partner’s ratings for the other eight. Accuracy was scored as whether a recommender’s predictions matched the relative ranking of their target’s ratings, across all 28 possible pairwise combinations of two jokes. We find that even with personal and contextual information about their target, the humans (M=56.8%, SE=1.4%) were still less accurate than the algorithms (M=62.4%, SE=1.2%).

Study 3 was very similar to Study 2—people who knew one another well were recruited in pairs (N=210), with the added feature that some human recommenders were able to see the machine’s prediction before making their own. Even humans with this seeming advantage (M=58.3%, SE=1.4%) were not much improved from baseline (M=58.0%, SE=1.4%) and still less accurate than the algorithms (M=62.8%, SE=1.0%), and this pattern held whether participants were asked to make recommendations for their partner, or a stranger pulled from a database. These results demonstrate that even though machines are better at making recommendations, humans were reluctant to rely on their judgment when producing them.

Study 4 compared humans and machines from the perspective of a recommendation consumer. We first asked subjects (N=100) to log onto Jester, an online collaborative filtering engine for jokes. They rated eight sample jokes and saw the first five recommended jokes from the algorithm. Then we asked them to compare the algorithm to someone they knew well, or an unknown other participant (between-subjects). In both cases, fewer people thought the algorithm would be more accurate (known=17%; unknown=46%) and did not prefer to receive algorithm recommendations over the human alternative (known=15%; unknown=37%).
In Study 5 we measured accuracy and consumer taste in the same experiment. Subjects (N=1300) first rated ten “menu” jokes, then were assigned to receive recommendations from either a machine or a human. They rated three sample jokes, then saw which three jokes their recommender had chosen for them. Again, the target had given higher ratings to the jokes the machine chose (M=2.8, SE=0.1) than the jokes the humans chose (M=2.5, SE=0.1) but when asked afterwards, participants gave higher ratings to the performance of human recommenders (M=4.1, SE=0.1) than machine recommenders (M=4.4, SE=0.1). These results confirm that resistance to machine recommenders persists even when the accuracy difference is made explicit.

These results shine a light on a rapidly developing facet of consumer choice. Consumer resistance to machine recommendations has been well-documented (e.g. Dawes 1979; Sunstein 2014). This raises important questions about how to get consumers to actually take recommendations from a machine, which would improve choices and reduce search costs. Future research will examine these barriers to resistance, and extend our findings to other domains.

REFERENCES