Reasoning About Diverse Evidence in Preference Predictions

Rachel Meng (rm3081@columbia.edu)
Columbia University, 3022 Broadway
New York, NY 10027 USA

Stephanie Y. Chen (Stephanie.Chen@chicagobooth.edu)
University of Chicago, 5807 S. Woodlawn Avenue
Chicago, IL 60637 USA

Daniel M. Bartels (bartels@uchicago.edu)
University of Chicago, 5807 S. Woodlawn Avenue
Chicago, IL 60637 USA

Abstract

People often incorporate the opinions of others to make predictions about the world, including their preferences for novel experiences and items. In two experiments, we explored how people use the opinions of dissimilar others in making such predictions. While social cognition research has found that similar others tend to influence our judgments more than dissimilar others, the diversity principle from category-based induction argues that we value evidence from diverse sources. Our results suggest that people seek and use information from dissimilar others differently when predicting their own preferences than when making predictions with more verifiable values. For self-relevant predictions, participants were less likely to seek the opinion of dissimilar advisors (Experiment 1) and more likely to contrast their judgments away from these advisors' opinions (Experiment 2).

Keywords: Advice; category-based induction; diversity; preferences; social influence.

Introduction

We frequently use the opinions of others as a basis for inductive reasoning, including when making predictions about our own preferences. Consider a scenario where a person wants to decide whether or not to see a new movie she knows very little about. To predict how much she will enjoy this unfamiliar movie, she can solicit opinions from others who have already seen it. But whose advice does she value more—that of people with a wide variety of movie tastes, or that of only people with movie tastes like her own? The current research asks whether we seek the opinions of individuals who are similar or dissimilar to us as well as how we use these opinions to inform predictions.

Work in the social cognition literature has found that similar others tend to be more influential in our judgments than dissimilar others (Festinger, 1954; Heider, 1958; Suls, Martin, & Wheeler, 2002). Not only do people treat similar others as reliable and attractive sources of information, particularly when they can easily discriminate their own tastes in a domain (Yaniv, Choshen-Hillel, & Milyavsky, 2011), but they tend to discount advice from those less like them (Twyman, Harvey, & Harries, 2008). From a social influence perspective, we might expect that a person who is predicting her own tastes would prefer to *assimilate* her opinions to those of similar others and *contrast* her opinions away from (or ignore) those of dissimilar others.

However, there is reason to believe that people may value dissimilar others in these settings. The results of a pilot study suggest that participants positively weighed the judgments from diverse advisors when inferring their own utility for a novel stimulus. In this study, 156 Amazon Mechanical Turk (MTurk) respondents viewed evaluations of an unfamiliar movie from a pair of movie-goers, one who was similar to them and another who was dissimilar.[1] Both movie-goers rated the movie very highly. We asked participants to describe how they would use this information to predict how much they would like the target movie.

Responses revealed that the dissimilar movie-goer's opinion informed most people's predictions. The majority of participants (61%) indicated that the dissimilar movie-goer's positive rating strengthened the likelihood that they would also enjoy the movie.[2] A significantly smaller proportion, 21% ($\chi^2(1) = 34.7$, $p < .001$), reported attending only to the similar person's rating.[3] These results support the idea that congruent opinions expressed by a dissimilar (and more diverse) set of individuals can favorably influence preference predictions.

Reasoning About Diverse Evidence

The findings from the pilot study are in line with the *diversity principle* discussed in category-based induction, according to which evidence from diverse sources support

[1] We manipulated perceived (dis)similarity by informing participants that the movie-goers agreed with them 80% versus 20% of the time in a movie evaluation task. A pretest ($N = 101$) confirmed that an advisor pair with an (80%, 20%) overlap in preferences to the participant was perceived as less similar *to each other* compared to a pair with an (80%, 80%) overlap ($M_{(80\%,20\%)} = 2.88$ ($SD = 1.37$), $M_{(80\%,80\%)} = 6.07$ (.97), $t(99) = 13.7$, $p < .001$).

[2] Sample response: "[T]hat both individuals gave the movie a strong rating, as well as two people with both similar AND different tastes from mine, affirms to me the movie is probably well liked by all and that I am likely going to enjoy it."

[3] Sample response: "I take my cue from how similar people's tastes are to mine. I am fairly picky and so if people like similar things it's a pretty reliable cue to take."

stronger arguments and broader generalizations than evidence from less diverse sources (e.g., Heit, 2000). These diverse samples create a stronger basis for generalization because they better cover the category of interest (Osherson et al., 1990). For example, Osherson and colleagues (1990) found that people judged arguments to be stronger when supported by diverse premises, both when the conclusion category was general (where the conclusion category is superordinate to the premise categories; e.g., generalizing from lions and goats to *all mammals*) and specific (where the conclusion and premise categories exist at the same level of the conceptual hierarchy; e.g., generalizing from lions and goats to *giraffes*). The specific case is analogous to our example scenario where someone is inducing her own movie preferences from those of other movie-goers, as the conclusion category (the self) and the premise category (another movie-goer) lie at the same level of specificity.

People are also sensitive to premise diversity when searching for information to support inferences. In these tasks, they typically prefer to seek diverse, rather than similar, pieces of evidence when judging the validity of generalizations (López, 1995; Rhodes, Brickman, & Gelman, 2008). For example, when assessing whether a blank, or unfamiliar, property (e.g., *has sesamoid bones*) holds for all mammals, participants would rather test whether it holds for lions and goats than for lions and leopards (López, 1995).

Predicting preferences based on others' recommendations departs from category-based induction tasks in at least two substantial ways. First, the types of predictions under scrutiny in this paper are inherently social, whereas many category-based induction tasks have focused on whether we prefer information or items from diverse biological categories and locations. Second, preference predictions implicate matters of taste (i.e., subjective predicates such as *liking a movie*) rather than facts about the world.

Despite these differences, the diversity principle may nevertheless apply in these contexts. Knowing the opinions of people similar *and* dissimilar to you can lead to strong predictions. If a pair of advisors with tastes both like and unlike your own both enjoyed a movie, you might reasonably infer that you would enjoy it, too—perhaps even more so than if a pair of advisors with only similar tastes enjoyed the movie, because the position is more broadly supported. When people with divergent tastes agree that a movie is good, you might conclude that it is more likely to be universally liked; hence, you will like it as well.

Previous work in social cognition and advice taking suggest that the opinions of dissimilar others can be more influential for more "objective," verifiable judgments (Goethals & Nelson, 1973) and for judgments about others' actions (Gino, Shang, & Croson, 2009) compared to ones related to our personal values and behavior. Thus, the current research explores how people select and use others' opinions when making predictions about their own, subjective preferences versus ones that take on more verifiable values. This comparison will allow us to examine whether ideas from the diversity principle differentially apply to these two situations.

Experiment 1: Selecting Evidence

The pilot suggested that people may positively incorporate the opinions of dissimilar others when making inferences about their own tastes. Experiment 1 examined whether these explicit self-reports matched how people actually seek the opinions of others to make preference predictions. Participants completed an evidence selection task in which they were asked to solicit opinions from a panel of "regular movie-goers" (reviewers, or advisors; we use these terms interchangeably) in order to predict how much they would like an unfamiliar movie. Panel reviewers were described to differ in how much their movie preferences overlapped with those of the participant; this perceived variation in similarity allowed us to assess whether people preferred solicit advice from similar or dissimilar others.

We contrasted these judgments about personal preferences (e.g., "how much will I like a movie?") against more verifiable beliefs about the movie. Specifically, our design used three such judgments: (a) how much *the average person* would like the movie, (b) how *critically acclaimed* the movie would be, and (c) how *successful* the movie would be at the box office (in terms of money made). This comparison allowed us to test whether people are more likely to sample similar (vs. dissimilar) others when making self-relevant predictions relative to predictions involving more verifiable values (items a, b, and c above).

Method

Design We randomly assigned participants to one of four between-subjects conditions (prediction condition: self, average person, critical acclaim, box office success).

Participants Two hundred and one U.S. respondents from MTurk completed an online study in return for $1.

Procedure We informed participants that they were tasked with making a prediction about a new, "mystery" movie and had the opportunity to learn the opinions of other reviewers who have seen it. In part one, participants indicated their preferences for different movie genres (e.g., comedies, science fiction, musicals) from 1 *(I hate this genre most of the time*) to 7 (*I love this genre most of the time*).

In part two, we presented all participants with the same stimuli: a panel of 12 anonymous reviewers described as regular movie-goers, ranked by their similarity to the participant's preferences based on their previous genre ratings. Each reviewer was labeled with an overlap rank (e.g., "Reviewer 358 [overlap rank: 1]"), with lower ranks signifying greater overlap with the participant's movie preferences. Participants read that each reviewer on the panel had seen and rated the movie, and were asked to choose the three reviewers whose opinions they most prefer to seek in order to make the prediction corresponding to their condition (see Table 1).

Table 1: Prediction conditions, Experiments 1 and 2.

Prediction condition	Evidence selection task wording: *Whose opinion would you like to get in order to predict…*
Self	*…how much you would like the movie?*
Average person	*…how much the average person would like the movie?*
Critical acclaim	*…how critically acclaimed the movie is likely to be?*
Box office success	*…how successful the movie is likely to be at the box office (i.e., money made)?*

Data Analysis To examine participants' selection patterns, we calculated two scores for each individual that quantified their selection of dissimilar advisors. First, we computed a *dissimilarity mean score* (DMS), given by the simple average of the overlap ranks of the three chosen advisors. Since lower ranks correspond to greater overlap in movie preferences with the participant, higher dissimilarity mean scores would indicate selection of more dissimilar advisors.

Second, we calculated a *dissimilarity spread score* (DSS), which indicated how widely and evenly a participant's selected ranks were spread across the panel of 12 reviewers. This score was obtained by computing the pairwise differences in ranks for each selection, taking the absolute value of these differences, and summing their minima across the three selections. More generally:

$$DSS_s = \sum_{i=1}^{n} \min_j |x_j - x_i|, \ \forall i,j \in [1,n], i \neq j, n \geq 2, \qquad (1)$$

where s indexes the participant, n is the total number of selected advisors (in Experiment 1, n = 3), i and j are any two unique selected advisors, and x_i and x_j are the respective ranks of advisors i and j.

For example, if a participant selected advisors with overlap ranks of 1, 6, and 12, the calculation would proceed as follows: For the first advisor, perform these two subtractions: $|1-6| = 5$, $|1-12| = 11$. For the second advisor: $|6-1| = 5$, $|6-12| = 6$. For the third: $|12-1| = 11$, $|12-6| = 6$. The resulting score is then the sum of the minimum of each of these differences: $5+5+6 = 16$.[4]

The spread score, while correlated with the mean score, confers additional insight into how broadly participants selected advisors. A participant who chose advisors with ranks $\{5,6,7\}$ would have the same DMS as one who chose $\{1,6,11\}$, but the second participant would have sampled a wider range of advisors (composed of a very similar, moderately similar, and very dissimilar advisor).

Strictly speaking, as we only revealed information about how similar the advisors were to the participant, we cannot know how similar they were to each other. Thus, the DSS is distinct from the concept of premise diversity in the category-induction literature, which states that premises that are less similar to each other provide stronger support for generalizations to the extent that they cover a given category (Osherson et al., 1990). Some evidence, however, suggests that participants may nevertheless perceive these concepts to be equivalent in our paradigm. According to a pretest (N = 119) conducted on a separate sample from the same pool, respondents believed that the greater the distance between two reviewers' overlap ranks, the more dissimilar they were to each other. In fact, judgments of inter-advisor dissimilarity increased linearly with rank distance ($F(1,118) = 290$, $p < .001$).[5] To the extent that our participants perceived advisors with a greater spread in ranks as more dissimilar from each other (relative to those with a smaller spread), the DSS defined in Equation 1 is likely to converge with notions of evidential diversity.

Results and Discussion

A univariate ANOVA on the two measures defined above revealed that participants' choices for their top three advisors differed across conditions (see Figure 1), both in terms of dissimilarity mean scores (M_{self} = 3.67 (1.94), $M_{average}$ = 4.96 (2.27), $M_{acclaim}$ = 4.46 (2.43), $M_{success}$ = 5.03 (2.31), $F(3,197) = 4.22$, $p = .006$) and dissimilarity spread scores (M_{self} = 7.00 (5.03), $M_{average}$ = 9.39 (5.55), $M_{acclaim}$ = 8.80 (5.78), $M_{success}$ = 9.89 (5.83), $F(3,197) = 2.79$, $p = .04$).

The results of a planned contrast found that, compared to the self condition, participants preferred to sample significantly more dissimilar advisors in the three other conditions ($M_{non\text{-}self}$ = 4.82 (2.34), $t(197) = 3.25$, $p = .001$). In particular, compared to the self condition, dissimilarity mean scores were higher for predictions about the average person ($t(197) = 2.95$, $p = .004$), box office success ($t(197) = 3.16$, $p = .002$), and, to a lesser extent, critical acclaim ($t(197) = 1.76$, $p = .08$). The same pattern was obtained for DSS. Compared to the self condition, participants preferred to sample a broader set of advisors in the three other conditions ($M_{non\text{-}self}$ = 9.36 (5.72), $t(197) = 2.69$, $p = .008$). More specifically, spread scores were higher for the average person ($t(197) = 2.19$, $p = .03$), success ($t(197) = 2.71$, $p = .007$), and, to a lesser extent, critical acclaim ($t(197) = 1.60$, $p = .11$).

In sum, participants solicited opinions more broadly when predicting more verifiable features of the movie than when predicting their own preferences. Approximately half (51%)

[4] Unlike variance or standard deviation, the DSS takes into account the relative distances among all selected ranks and is greater when they are more *evenly* dispersed (e.g., {1, 6, 12}) as opposed to extreme (e.g., {1, 11, 12}).

[5] For example, participants inferred an advisor pair with a rank distance of 11 (1 vs. 12) to be much less similar to each other (*1 = very dissimilar; 7 = very similar*) compared to a pair with rank distance 5 (e.g., 1 vs. 6; $M_{\{1,12\}}$ = 2.03 (1.51), $M_{\{1,6\}}$ = 3.62 (.91), $t(118) = -10.6$, $p < .001$). This pair was in turn judged less similar than a pair with rank distance 1 (e.g., 1 vs. 2; $M_{\{1,2\}}$ = 5.72 (1.44), $t(118) = -17.0$, $p < .001$). The same judgments held for an advisor triad with ranks $\{1,6,11\}$ versus one with ranks $\{5,6,7\}$ ($M_{\{1,6,11\}}$ = 3.04 (1.24), $M_{\{5,6,7\}}$ = 5.31 (1.15), $t(118) = -13.2$, $p < .001$).

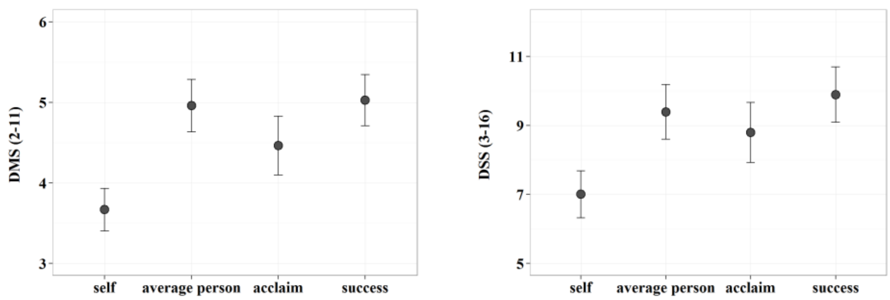

Figure 1: Dissimilarity mean (*left*) and spread scores (*right*). Error bars are standard errors (+/- SE).

in the self condition selected the three most similar advisors (ranks 1, 2, and 3), compared to 27%, 25%, and 39% in the average person, success, and critical acclaim conditions. Moreover, whereas only 44% selected an advisor with a rank greater than 6 in the self condition (higher ranks indicate greater dissimilarity), 67%, 68%, and 55% did so in the average person, success, and critical acclaim conditions.

Unexpectedly, choice patterns in the critical acclaim condition were not as reliably different from the self condition as they were in the average person and success conditions. We speculate that this may be due to a "better than average" effect (Alicke, 1985) where people may believe they have above-average taste in movies, not unlike that of movie critics. Accordingly, participants may have treated predicting a movie's critical acclaim as more similar to predicting their own preferences.

Experiment 2: Updating Predictions

The results of Experiment 1 suggest that people are more likely to seek opinions from dissimilar advisors when forming more verifiable judgments (e.g., a movie's box office success) than when predicting their own preferences. A separate but related question is how they then use the opinions of dissimilar advisors to inform such predictions.

In Experiment 2, we examined how people use opinions from similar versus dissimilar reviewers to update their predictions (the four predictions used in Experiment 1). Participants saw ratings from two different advisors sequentially, making one prediction after seeing each advisor's opinion. Half the participants saw two advisors who were similar to them in terms of movie preference (the similar pair condition), while the other half saw one similar and one dissimilar advisor (the dissimilar pair condition). In both conditions, the two advisors rated the movie highly.

As the similar advisor was always presented first, participants should rate the movie positively after seeing that advisor's rating. The critical question is how participants updated their initial prediction upon learning of a second, dissimilar advisor's positive rating. The diversity principle argues that convergent evidence from diverse sources should strengthen inductive inferences. Assimilation of the dissimilar other's opinion—manifested as a more positive prediction about the movie—would be consistent with such a diversity effect: If a wider range of individuals both recommend a movie, then it must be good (because it likely appeals to a broad audience). Counter to the diversity principle, contrast from the second, dissimilar other's opinion—manifested as a less positive prediction about the movie—would imply a strategy that takes the diverse preferences of others as a negative cue.

Method

Design We randomly assigned participants to one of eight conditions in a 4 (prediction condition: self, average person, critical acclaim, box office success) × 2 (advisor pair: similar, dissimilar) between-subjects design.

Participants Three hundred and ninety-eight U.S. MTurk respondents completed an online survey in return for $1.

Procedure Part one of Experiment 2 was identical to Experiment 1. In part two, participants read the same information about the panel of 12 reviewers as in Experiment 1. They were then told that they would see the ratings of two reviewers from this panel and make the prediction corresponding to their prediction condition (Table 1) after seeing each reviewer's rating of the movie.

Participants in all conditions viewed the same rating information. They first learned that the most similar reviewer (with an overlap rank of 1) rated the target movie a 9 out of 10 (*1 = disliked very much*; *10 = liked very much*). Participants next rated the movie on the dimension that corresponded to their prediction condition based on this first advisor (e.g., average person: "How much do you think the *average person* would like this movie?"). All predictions were rated on the same 10-point scale, where higher ratings indicate more favorable predictions about the target movie.

After their initial prediction, participants learned about a second reviewer. In the similar pair condition, participants saw the rating of the second most similar reviewer to themselves (rank 2). In the dissimilar pair condition, they saw the rating of the most dissimilar reviewer (rank 12). In both advisor pair conditions, the second reviewer also gave the movie a positive rating (8.5 out of 10). After viewing this second opinion, participants made another prediction, identical in format to the first.

Data Analysis To measure the direction and magnitude of belief revision in response to the second advisor's opinion, we subtracted each participant's first prediction from their second to form a *difference score*. A score of zero would mean an individual did not change her initial prediction, while positive and negative values would mean that she rated the movie more favorably and less favorably, respectively. Since this second advisor always agreed with the first, positive difference scores would indicate assimilation toward the second advisor's opinion, while negative difference scores would indicate contrast.

Results and Discussion

A 4 (prediction condition: self, average person, critical acclaim, box office success) × 2 (advisor pair: similar, dissimilar) factorial ANOVA on difference scores found main effects of each factor (prediction condition: $F(3,390) = 8.71$, $p < .001$; pair: $F(1,390) = 16.3$, $p < .001$), qualified by an interaction ($F(3,390) = 6.85$, $p < .001$).

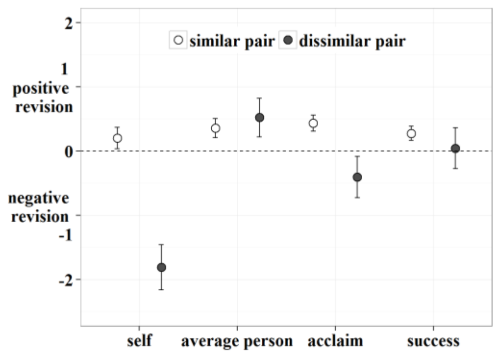

Figure 2: Difference scores by prediction condition and advisor pair. Error bars are standard errors (+/- SE). The dotted line at the zero intercept indicates no revision.

As shown in Figure 2, the results of the self condition differed from those of the other conditions. An analysis of the simple effects of advisor pair on difference scores within each prediction condition revealed that when predicting for the self, seeing a similar versus dissimilar pair differentially influenced final judgments ($F(1,390) = 30.2$, $p < .001$). The mean for the dissimilar pair condition was significantly less than zero (M = -1.81 (2.52), $t(51) = -5.17$, $p < .001$), suggesting that these participants contrasted their opinions away from the dissimilar advisor. This is the only condition where contrast effects surfaced. In the similar condition, the mean did not differ from zero (M = .20 (1.12), $t(44) = 1.20$, *ns*), suggesting that the second similar advisor's opinion conferred no marginal value.[6]

In the critical acclaim condition, as with the self condition, the simple effects of advisor pair on difference scores revealed that seeing a similar versus dissimilar pair differentially influenced final judgments ($F(1,390) = 5.56$, $p = .019$). However, the impact of the similar and dissimilar advisor was quite different here than in the self condition. Participants made stronger (i.e., more positive) predictions about the movie after seeing a second, similar advisor's opinion. One-sample *t*-tests against zero found that participants positively adjusted their initial beliefs when predicting critical acclaim (M = .43 (.83), $t(45) = 3.54$, $p = .001$). In contrast, when the second reviewer was dissimilar, participants did not change their initial predictions (M = -.40 (2.43), $t(56) = -1.25$, *ns*).

For predictions about the average person and box office success, no differences emerged in the amount of belief updating between the similar and dissimilar advisor pairs (Fs < 1). The similar advisor's positive opinion of the movie lead to upward belief revisions, both for the average person (M = .36 (1.10), $t(55) = 2.42$, $p = .019$) and success (M = .28 (.74), $t(46) = 2.55$, $p = .014$). When the second advisor was dissimilar, there was no evidence of a contrast effect for either condition. If anything, people tended to *assimilate* the opinion of dissimilar others when predicting for the average person, as evidenced by a directional upward revision (M = .52 (2.07), $t(47) = 1.74$, $p = .088$).

Consistent with the results of Experiment 1, these results revealed that participants' use of the opinions of dissimilar others differed in the self condition versus in the other three conditions. Specifically, participants contrasted their own predictions away from the opinion of a second, dissimilar advisor.[7] No evidence of such a contrast effect emerged for predictions pertaining to the average person and box office success. Thus, preference diversity appears to be treated as a negative cue only for judgments of personal preference.

Interestingly, in Experiment 2 participants seemed to find evidence from a dissimilar advisor most informative in the self condition (i.e., these participants revised their initial predictions the most). The results of Experiment 1, however, suggest that people were also *least* likely to seek out the opinion of a dissimilar other in this condition—the very condition where it may be *most* informative.

General Discussion

The opinions of others often shape, or even serve as the basis for, our own beliefs about the world. These beliefs can influence both our inferences (e.g., "Will I like that movie?") and our choices (e.g., "Will I go see it?").

Our experiments used ideas from category-based induction, social cognition, and advice taking/seeking to explore how and when information from diverse others is used in inductive reasoning. The results suggest that people value similar and dissimilar others' opinions differently when predicting their own preferences for a novel stimulus. For this type of self-relevant judgment, participants were less likely to sample information from dissimilar advisors (Experiment 1) and contrasted their predictions away from the opinion of dissimilar advisors—even when a similar advisor expressed a congruent opinion (Experiment 2). We have found the same tendency to contrast predictions from dissimilar others among university students and in other domains (e.g., music, restaurants). Therefore, these patterns

[6] This is likely not due to a ceiling effect, as the average initial prediction in this condition was 7.6 out of 10, comparable with the other three prediction conditions where there is upward revision.

[7] We observed similar results in other designs that manipulated the order of the similar and dissimilar reviewer, as well as when reviews were presented simultaneously.

appear to hold across different populations and kinds of preference predictions.

The results we have presented are somewhat at odds with findings in the category-based induction literature that have documented the greater appeal and informativeness of diverse evidence. Why, then, do our participants—despite their explicit self-reports in the pilot—not seem to value sampling the opinions of people with a wide range of preferences when making self-relevant judgments? A category-based induction perspective offers two possible explanations. First, participants may believe that they belong to a subordinate category of movie watchers and thus, to determine their preference for a given movie, only the opinions from people who belong to that category (e.g., science fiction aficionados) are relevant. In this case, people may still appreciate diversity—but only to the extent they perceive diverse advisors to suitably cover the subordinate category of which they see themselves as members.

Second, participants may believe that individuals with similar tastes to their own share features which cause them to like the same movies. The relevance theory of induction (Medin et al., 2003) proposes that we are often sensitive to the causal scenarios and common properties between the premises and the conclusion. This would suggest that how we seek information from others on matters of taste depends on what we believe causes our preferences and whether we think these causes are present in others. In our paradigm, people may believe there their movie preferences and those of similar others stem from the same cause (e.g., an affinity for indie movies). One can easily imagine how sharing a similar taste in movie genres would lead to similar movie preferences in general. Had the advisors in our panel overlapped with participants on features bearing no causal relationship to liking movies (e.g., they had eaten the same food for breakfast), participants would likely not have been as inclined to seek the opinions of "similar" others. Further work is necessary to discern the role of these two possibilities in shaping preference predictions.

Importantly, the results of Experiment 1 also suggest conditions when people *are* more sensitive to evidential diversity. Relative to predictions about their own taste, people were more likely to sample both similar and dissimilar reviewers when predicting the preferences of the average person and a movie's box office success. What explains this difference? One possibility is that in the average person and success conditions, the conclusion "all people will like the movie" is a useful proxy for the target prediction. Knowing the opinion of both dissimilar and similar others (relative to knowing the opinion of only similar others) produces greater coverage of the general conclusion category "all people" and, consequently, supports stronger predictions.

Taken together, these results have implications for what types of information people seek when making inferences in different contexts. For example, if we were tasked with judging the quality of a CogSci paper for a review, we may prefer to poll conference attendees with both similar and dissimilar interests to our own. On the other hand, if we wanted to determine which paper we should spend our time reading for personal pleasure, we may prefer to only poll colleagues who share our taste in papers. Identifying precisely which factors affect how broadly we sample advice poses another important topic of future research.

While research in category-based induction has revealed a great deal about the induction process, much of it often focuses on biological and artificial categories. However, this same process is frequently at work in preference-rich and social situations like the ones we have explored in this paper. Testing how people sample and use the opinions of others in these settings, including when diversity effects are likely to prevail, brings us closer to understanding how individuals reason about evidential diversity "in the wild."

References

Alicke, M. D. (1985). Global self-evaluation as determined by the desirability and controllability of trait adjectives. *Journal of Personality and Social Psychology, 49,* 1621-1630.

Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, *7*, 117-140.

Gino, F., Shang, J., & Croson, R. (2009). The impact of information from similar or different advisors on judgment. *Organizational Behavior and Human Decision Processes*, *108*, 287-302.

Goethals, G. R. & Nelson, R. E. (1973). Similarity in the influence process: The belief-value distinction. *Journal of Personality and Social Psychology*, *25*, 117-122.

Heider, F. (1958). *The psychology of interpersonal relations.* Hoboken, NJ: John Wiley & Sons Inc.

Heit, E. (2000). Properties of inductive reasoning. *Psychonomic Bulletin & Review, 7*, 569-592.

López, A. (1995). The diversity principle in the testing of arguments. *Memory & Cognition*, *23*, 374-382.

Medin, D. L., Coley, J. D., Storms, G., & Hayes, B. K. (2003). A relevance theory of induction. *Psychonomic Bulletin & Review*, *10*, 517-532.

Osherson, D. N., Smith, E. E., Wilkie, O., López, A., & Shafir, E. (1990). Category-based induction. *Psychological Review*, *97*, 185-200.

Rhodes, M., Brickman, D., & Gelman, S. A. (2008). Sample diversity and premise typicality in inductive reasoning: Evidence for developmental change. *Cognition, 108*, 543-556.

Suls, J., Martin, R., Wheeler, L. (2002). Social comparison: Why, with whom, and with what effect? *Current Directions in Psychological Science*, *11*, 159-163.

Twyman, M., Harvey, N., & Harries, C. (2008). Trust in motives, trust in competence: Separate factors determining the effectiveness of risk communication. *Judgment and Decision Making*, *3*, 111–120.

Yaniv, I., Choshen-Hillel, S., & Milyavsky, M. (2011). Receiving advice on matters of taste: Similarity, majority influence, and taste discrimination. *Organizational Behavior and Human Decision Processes*, *115*, 111-120.