# Journal of Neuroscience, Psychology, and Economics


## Gaze-Informed Modeling of Preference Learning and Prediction


Stephanie M. Smith and Ian Krajbich



Online First Publication, July 15, 2019. http://dx.doi.org/10.1037/npe0000107

American Psychological Association

Journal of Neuroscience, Psychology, and Economics


1937-321X/19/$12.00

http://dx.doi.org/10.1037/npe0000107

# Gaze-Informed Modeling of Preference Learning and Prediction

Stephanie M. Smith and Ian Krajbich
The Ohio State University

Learning other people's preferences is a basic skill required to function effectively in society. However, the process underlying this behavior has been left largely unstudied. Here we aimed to characterize this process, using eye-tracking and computational modeling to study people while they estimated another person's film preferences. In the first half of the study, subjects received immediate feedback after their guess, whereas in the second half, subjects were presented with four random first-half outcomes to aid them with their current estimation. From a variety of learning models, we identified two that best fit subjects' behavior and eye movements: k-nearest neighbor and beauty contest. These results indicate that although some people attempt to form a high-dimensional representation of other people's preferences, others simply go with the average opinion. These strategies can be distinguished by looking at a person's eye movements. The results also demonstrate subjects' ability to appropriately weight feedback in their estimates.

*Keywords:* preference estimation, computational modeling, eye-tracking, attention, individual differences

*Supplemental materials:* http://dx.doi.org/10.1037/npe0000107.supp



People often need to learn the preferences of their friends, family, and colleagues, so as to improve their recommendations, gift-giving abilities, and, more generally, their relationships. Estimating another's preferences can be a vital task, especially if one should need to make a decision on their behalf. Maintaining the individual's autonomy is heralded as the gold standard when surrogates make decisions for another (Minogue, 1996); to preserve autonomy, the surrogate must know the recipient's preferences. An important, unanswered question, though, is precisely how a person learns another's preferences.

For example, if you had to recommend a restaurant to someone you barely knew, what would you do? You might, for instance, start with your personal favorite place. Alternatively, you might suggest the general consensus among your friends. Then, after hearing this diner's evaluation of your idea, how would you adapt your next recommendation? How would you integrate their feedback with your own knowledge and opinions? Clearly, the process involved in learning another's preferences is quite nuanced and complex.

Computer scientists have been developing machine learning techniques for predicting people's preferences. For example, Netflix predicts movie preferences (and makes suggestions) based on the feedback it receives. It uses rating, viewing, scrolling, and search behavior as well as the similarities and differences between watched and unwatched films to determine the predicted rating for another film (and to calculate, ultimately, whether or not to suggest this

Stephanie M. Smith, Department of Psychology, The Ohio State University; Ian Krajbich, Departments of Psychology and Economics, The Ohio State University.

This work was supported by NSF Career Grant 1554837 (to Ian Krajbich) and NSF GRFP Grant DGE-1343012 (to Stephanie M. Smith). We thank R. Gwinn for sharing experiment code and A. Hedstrom for film advice. The data sets analyzed for this study can be found in the Open Science Framework (https://osf.io/k7rjz/?view_only=a9178b2938794f7cbd19d7bdf037598e).

Correspondence concerning this article should be addressed to Ian Krajbich, Departments of Psychology and Economics, The Ohio State University, 1827 Neil Avenue, Columbus, OH 43210. E-mail: krajbich.1@osu.edu

other film; Vanderbilt, 2013). There is substantial overlap between these machine-learning techniques and the models used in research on human learning, ranging from basic reinforcement learning (Sutton & Barto, 1998) to more complex models like k-nearest neighbor (Aha, Kibler, & Albert, 1991). Thus, these models provide possible frameworks for understanding human preference estimation. For example, people might expect others to have an underlying preference schematic—an internal, multidimensional map of films, for instance—where the similarity between any two films relates inversely to the difference in preference between them. If people do have such representations of their preferences, then others might try to learn (and use) this underlying structure.

At the same time, the existing research is less than encouraging when it comes to human competence in predicting others' preferences. For instance, some findings indicate that people predict others to be more risk-seeking than themselves for gambles over gains and losses (Hsee & Weber, 1997), whereas others show that people predict others' preferences to be closer to risk neutrality (Faro & Rottenstreich, 2006). In a more applied context, gift-givers often fail to give their recipients the preferred gift, even when the giver and recipient know each other very well (Givi & Galak, 2017). Perhaps the most alarming research demonstrates that when participants are asked to predict whether a family member would want life-sustaining treatment in a variety of health scenarios, their estimations were largely inaccurate and more closely mimicked the estimator's preferences than the recipient's (Fagerlin, Ditto, Danks, Houts, & Smucker, 2001). It is therefore also possible that in a context such as this, people do not learn or adapt, but instead use heuristics. For instance, in predicting the preferences of someone that they do not know well, a reasonable strategy might be to guess the *average* individual's preference. With some knowledge about the other person, one might instead use the average as the starting point in an anchor-and-adjustment process (Tamir & Mitchell, 2010).

It seems plausible then that humans might utilize some combination of learning techniques and simple heuristics. The question is which models most closely match peoples' actual choice processes?

Previous research has demonstrated the usefulness of eye-tracking and other process-tracing data for inferring choice processes (Aimone, Ball, & King-Casas, 2016; Ashby, Dickert, & Glöckner, 2012; Gharib, Mier, Adolphs, & Shimojo, 2015; Johnson, Camerer, Sen, & Rymon, 2002; Kim, Seligman, & Kable, 2012; Knoepfle, Wang, & Camerer, 2009; Konovalov & Krajbich, 2016; Lindner et al., 2014; Lohse & Johnson, 1996; Polonio, Di Guida, & Coricelli, 2014; Russo & Leclerc, 1994; Venkatraman, Payne, & Huettel, 2014). In particular, this work has demonstrated that information acquisition, as indexed by looking, has a significant impact on people's choices (Cavanagh, Wiecki, Kochar, & Frank, 2014; Fiedler & Glöckner, 2012; Fisher, 2017; Krajbich, Armel, & Rangel, 2010; Krajbich, Lu, Camerer, & Rangel, 2012; Krajbich & Smith, 2015; Milosavljevic, Navalpakkam, Koch, & Rangel, 2012; Orquin & Mueller Loose, 2013; Pärnamets et al., 2015; Shimojo, Simion, Shimojo, & Scheier, 2003; Smith & Krajbich, 2018, 2019; Stewart, Hermens, Matthews, 2015; Towal, Mormann, & Koch, 2013; Vaidya & Fellows, 2015). Specifically, when subjects spend longer gazing at an option, they gather more evidence about said option and are subsequently more likely to choose it from a set of alternatives; this finding holds even in perceptual judgments (Tavares, Perona, & Rangel, 2017).

In this study, we sought to combine computational modeling with eye-tracking data to study the preference estimation process. To this end, we studied human subjects as they attempted to guess a passive subject's values for a variety of movies. The experiment consisted of two blocks. In the first block, subjects made predictions for 100 movies, receiving feedback about the true value after each guess. In the second block, subjects made prediction for 100 more movies, this time without feedback between trials. However, here we provided subjects with onscreen feedback about prior guesses and true values. We tracked their eye-movements while they inspected this information and incorporated it into their predictions. The aim of the study was to use the choice data from the first block of trials to identify the best-fitting learning model (out of our candidate set) and then to validate this classification and characterize the different learning processes using the eye-tracking data in the second block.

We find that different subjects do seem to use different strategies. With the choice data, we show that some subjects mimicked a simple, static heuristic to make their estimates, whereas others used a dynamic strategy. With the eye-tracking data, we see that our subjects used explicitly provided feedback to update their preference estimations in different ways, depending on their choice-based classification.

## Method

### Subjects

Thirty-six university students participated in this study: one as a passive subject and 35 as active subjects. This study was approved by The Ohio State University Human Subjects Internal Review Board, and all subjects gave informed written consent before participating.

### Materials

Stimuli were presented using the MATLAB (MathWorks, 2014) Psychophysics Toolbox (Brainard, 1997; Kleiner et al., 2007; Pelli, 1997). An EyeLink 1000 Plus was used to collect eye-tracking data. Attentional areas of interest were defined a priori, each containing a piece of information on the screen. Subjects indicated all of their responses using a standard computer keyboard and mouse.

### Procedure

The study design was adapted from a previous paradigm (Janowski, Camerer, & Rangel, 2013). One recipient (the passive subject) provided willingness-to-pay (WTP) ratings on a discrete scale of \$1 to \$10 for 200 films that he had previously seen. A list of these films can be found in the supporting material. Each active subject ($N = 35$) completed three tasks (Figure 1). In the first block (Figure 1a), subjects were informed that the passive subject had provided his WTP for several films and were then provided incentivized estimates about the passive subject's WTP for each of 100 films. The closer a subject's guess to the actual WTP, the more cash they earned (up to \$0.10 per guess, decreasing by \$0.01 for each unit increase in distance between predicted and actual WTP). Subjects received immediate feedback (the actual WTP) after each estimate.

In the second block, subjects estimated the passive subject's WTP for the remaining 100 films from the sample (Figure 1b). These estimates were also incentivized with the same payoff structure as the first block (up to \$0.10 per estimate), but subjects did not receive any feedback. However, while making each estimates, subjects were provided with four randomly selected first-block films, along with the associated feedback (guess and WTP) from each film. During this block, we tracked subjects' eye movements so as to investigate how a subject's gaze patterns might influence their estimation process. The sets of films in the first and second blocks were randomly determined at the subject level.

In the last task (the film characteristics task), subjects rated each of the 200 films on four spectrums: action versus nonaction, realistic versus fantasy, serious versus comedy, scary versus nonscary (Figure 1c). With ratings of each film on each of these four dimensions, we were able to transform these ratings into a measure of subjective similarity between each pair of films (based on four-dimensional Euclidean distance). Thus, more similar films (i.e., films that were rated more similarly on the four dimensions) would be closer together, in terms of Euclidean distance. Subjects also indicated their confidence in these four ratings on a separate scale.

## Results

Unsurprisingly, subjects were able to learn the preferences of our passive subject through explicit feedback. Indeed, subjects' predictions improved throughout the first block. The population correlation between absolute error (i.e., $| \, Guess - WTP \, |$) and trial was $-0.244$ ($p = .014$), indicating that absolute errors decreased as trial number increased. More details on subject performance and response times are available in the online supplemental materials.

### First Block Sorting

To better understand the process underlying the learning process, subjects' estimates in the first block were compared to a variety of learning models. Although our list is by no means

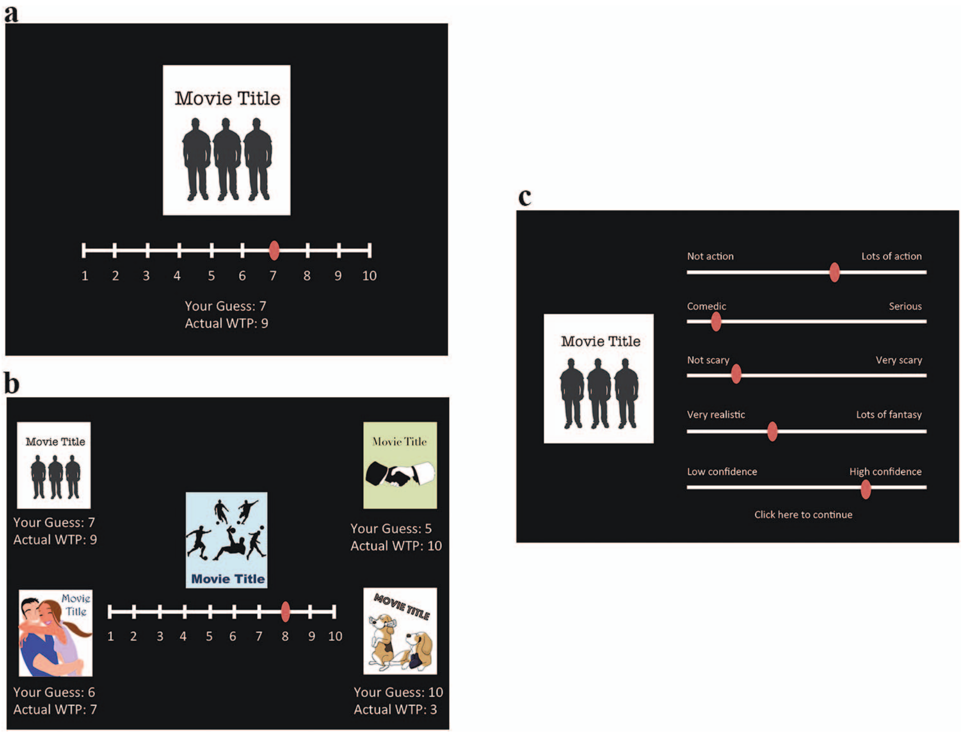


*Figure 1.* Experiment design. (a) First block: Subjects ($N = 35$) estimated a passive subject's willingness-to-pay for 100 films and were provided with immediate feedback. Estimates were incentivized for accuracy. (b) Second block: Subjects estimated the passive subject's willingness-to-pay for 100 new films and received no feedback, but, again, accuracy was incentivized. In addition, during each estimate, feedback from four randomly selected first-block films appeared onscreen. Subjects were eye-tracked during this block. (c) Last, subjects provided ratings on each of the 200 films on four spectrums (action, comedy, scary, and fantasy) and indicated their confidence in these judgments. See the online article for the color version of this figure.

exhaustive (for the sake of feasibility), we included a variety of well-known models from machine learning and decision psychology for comparison: reinforcement learning (RL), naïve learning (NL), anchoring and adjustment (AA), linear discriminant analysis (LDA), k-nearest neighbor (KNN), and beauty contest (BC).

**Reinforcement learning.** RL is a workhorse model that posits when people are rewarded for "good" behavior, they will consequently attempt to continue such behavior (Sutton & Barto, 1998). In this study, good behavior is that which minimizes the differences between a subject's estimate and the passive recipient's actual WTP. The model knows nothing about any given film, it simply adjusts its next guess upward or downward based on whether the previous guess was too low or too high.

$$RL: Guess_i = Guess_{i-1} + f * Error_{i-1}$$

$Error_{i-1}$ is the signed difference between the estimated WTP and actual WTP from the previous trial, which is integrated into the current estimate using a weight of $f$.

**Naïve learning.** Naïve subjects might, rather, than take into account the magnitude (and direction) of their errors, simply regurgitate the information that they have been given most recently. That is, they might take the arithmetic mean of the previous $n$ films' WTPs and use this average as their estimate.

$$NL: Guess_i = \frac{1}{n}\sum_{j=1}^{n} WTP_{i-j}$$

**Anchor and adjust.** AA (Tversky & Kahneman, 1975) is a heuristic for estimation in the judgment and decision-making literature. The proposed process involves "anchoring" to a potential estimate and then "adjusting" from said anchor to improve the estimate. Such adjustments from the anchor are often insufficient (Epley & Gilovich, 2006), resulting in imperfect estimates. The anchor on each trial is the previous trial's WTP; the adjustment amount $e$ is an individually fitted parameter. The model assumes that the subject knows in which direction to adjust the estimate.

$$AA: Guess_i = WTP_{i-1} \pm e$$

**Linear discriminant analysis.** LDA is a statistical classification technique that assumes that a set of multidimensional observations (in this case, films) come from different classes (in this case, WTPs). In this experiment, the subject-supplied ratings on each film provide an individual-level set of observations with which to construct a statistically optimal separation of the films into their respective categories. This model does not assume that subjects are updating their beliefs dynamically; it is a static model that assumes subjects have an underlying framework for how to estimate a film's WTP based on its spectrum ratings from the film characteristics task.

$$LDA: Guess_i = ar_i^1 + br_i^2 + cr_i^3 + dr_i^4$$

The $r_i^j$ values are the subjective ratings taken from each subject, whereas $a$, $b$, $c$, and $d$ shape the linear discrimination space created by the ratings.

**K-nearest neighbor.** KNN classification is another statistical method used in categorization. This strategy depends solely on an item's $k$ nearest neighbors. That is, the estimate for the current film is the average of the WTPs from the prior $k$ most similar films. Similarity is defined as the inverse four-dimensional Euclidean distance between any two films, according to each subject's ratings from the film characteristics task.

$$KNN: Guess_i = \frac{1}{k}\sum_{m=1}^{k} WTP_m$$

The KNN model is related to exemplar models of classification, which have been popularized in cognitive psychology (Nosofsky, 2011). For instance, the underlying idea behind the generalized context model (GCM; Nosofsky, 1986) is that when people attempt to categorize a new object, the judged similarity of this object to the existing exemplars of a given category determines whether (or not) it will be identified as part of that category. The GCM allows for context-dependent similarity judgments and allows different exemplars to be stored differently (i.e., stronger vs. weaker) in memory. In our model, we fix the similarity judgments to be a simple (inverse) Euclidean distance between the films and assume equal memory strength for each of the films. A much more complicated KNN model, for instance, could more closely mimic the GCM by taking into account how recently the subject encountered each of the films (to account for memory deterioration) in conjunction with the subjective similarity of each film. For the intents and purposes of this project, we opted to start with the simplest versions of each model, but this could be an interesting direction for future work.

**Beauty contest.** This model pays homage to the Keynesian beauty contest game from behavioral economics (Coricelli & Nagel, 2009; Keynes, 1936). In the traditional beauty contest game, the goal is to pick the most popular beauty contest entrant. That is, it is not necessarily optimal to pick one's own preference, but instead to (attempt to) choose the entrant believed to be picked the most by others. The extension to the current study is simple; when tasked with estimating someone else's film preferences, it is not necessarily optimal to offer one's personal opinion, but it might be useful to submit the opinion of an "average" person. The Internet Movie Database (IMDb.com) provides average ratings for films comprising huge numbers of voters, so these IMDb ratings served as a proxy for what subjects might expect the WTP of an "average" person to be.

$$BC: Guess_i = IMDb_i$$

$IMDb_i$ is the average rating given to the film on the Internet Movie Database (obtained from imdb.com). We tested an exact mapping between IMDb rating and subject guesses to limit the number of possible models and because the two metrics fall on the same scale (i.e., 1–10). However, as a useful extension, we also fit a version of the BC model with two free parameters: the slope and intercept for a regression of the subjects' guesses as a function of IMDb rating. More detail on this fitting procedure and the subsequent analyses is available in the online supplemental materials.

The optimal model for each subject's behavior in the first block was identified using a leave-one-out cross-validation process. One trial was left out on each iteration, and each model was fitted using the remaining trials; the best-fitting parameters for each model were used to predict the subject's estimate for the left-out trial. Thus, we obtained an out-of-sample prediction from each model for each trial and then compared the accuracy of all models in fitting each subject's actual guesses. The mean square error (*MSE*; average squared deviance between predicted and actual estimates for the subjects) of the fitted estimates was used as a goodness-of-fit metric to compare the models; the best-fitting model was the one with the lowest *MSE*.

An initial sort (based on minimal *MSE*) across these six strategies yielded four NL, 11 KNN, and 20 BC subjects. For the sake of pairwise comparison and meaningful group statistics, we recoded the four NL subjects into the remaining two groups, using the smaller of the two MSEs as the sorting mechanism, yielding 14 KNN and 21 BC. Subsequent results remain qualitatively unchanged if we instead exclude these four subjects.

Behaviorally, there were a few differences between the two sorted groups. Despite the fact that the passive subject's true WTPs were significantly positively correlated with the IMDb ratings ($r$ = .43, $p$ = $10^{-10}$; Figure 2), KNN subjects (mean absolute error = 2.09) performed slightly better in the first block than the BC subjects (mean absolute error = 2.26), $t$(25.036) = 2.135, $p$ = .043. In addition, as expected according to the static/dynamic nature of the BC/KNN models, BC subjects did not significantly improve over time (correlation between trial and average absolute error: $r$ = −0.15, $p$ = .15), whereas KNN subjects' performance did significantly improve (correlation between trial and average absolute error: $r$ = −0.22, $p$ = .03).

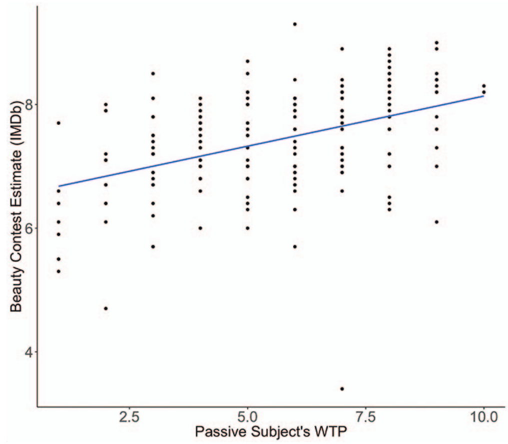


*Figure 2.* Relationship between the passive subject's true willingness-to-pay and the IMDb ratings. There is a significant positive relationship between the two ($r$ = .43, $p$ = $10^{-10}$), suggesting that beauty contest is a reasonable strategy to use in this task. The points are individual films and the blue line is a simple fitted regression through them. See the online article for the color version of this figure.

Surprisingly, there was no difference in performance in the second block, however (KNN: mean absolute error = 2.05; BC: mean absolute error = 2.10), $t$(28.403) = 0.514, $p$ = .611. There was no difference in response time between the two groups in either block (first block: $t$(15.165) = −1.339, $p$ = .200; second block: $t$(22.548) = −0.930, $p$ = .362).

In addition to providing the basis for our measure of subjective similarity, the film characteristics task also asked subjects to indicate how confident they were about their ratings of each film on the dimensions of action versus nonaction, realistic versus fantasy, serious versus comedy, and scary versus nonscary. If subjects were generally less confident in their ratings, then this suggests less familiarity with the films. We thought it might be possible that subjects who were less familiar with the films would be less capable of predicting how much an average person would like the film (i.e., using the BC strategy). However, we do not find substantial evidence for this prediction. A two-sample $t$ test of the average (subject-level) confidence across the two strategies did not provide evidence for a significant difference, $t$(24.6) = −0.9, $p$ = .37. A Kolmogorov–Smirnov test also did not reach significance, $D$ = 0.26, $p$ = .56. A histogram of the average confidence levels, separated by strategy, can be found in (Figure S3 in the online supplemental materials).

We do find a significant trial-level relationship between confidence and absolute error. Specifically, we find that when subjects were more confident about their film characteristic ratings, their guesses were closer to the true WTP of the passive subject (mixed effects regression with subject-level slope and intercept of absolute error on confidence, $\beta = -0.366$, $p < 10^{-7}$). We also fit a regression with an interaction between confidence and strategy, but neither the simple effect of strategy nor the interaction between strategy and confidence reached significance ($\beta = -0.101$, $p = .279$, and $\beta = -0.048$, $p = .722$, respectively).

## Strategy Differences in Gaze Patterns

After separating subjects into two distinct groups according to their behavior in the first block, we next examined how the gaze patterns differed between the two groups in the second block. One interesting difference between the two groups emerges in terms of which films they look at most. We might expect—based on the driving ideas of the strategy itself—that KNN subjects would look longer at more similar films (i.e., nearer neighbors), relative to less similar films, to better inform their estimates. On the other hand, we expect no such relationship for BC subjects because that strategy is not rooted in interfilm similarity. Indeed, this is precisely the pattern we see in the data.

For each subject, dwell time was regressed on the subjective similarity between the gazed-at film and the target film. A two-sample *t* test on the similarity regression coefficient indicates that similarity is a significantly stronger predictor of dwell time for KNN subjects than for BC subjects, $t(23.728) = 2.368$, $p = .026$. When we extend this analysis to include dwell times on the WTPs and guesses (in addition to just the posters), the difference is not quite as strong, $t(19.529) = 1.763$, $p = .094$. No significant difference in the relationship to similarity was found for dwell times on the WTPs or guesses alone.

When we rerun the first regression as a multilevel model with an interaction term (as well as random intercepts for subjects and films), we see a marginal positive interaction between strategy and similarity, $\beta = 0.048$, $p = .089$. Rerunning the second regression (that includes the WTPs and guesses) yields a similar result: $\beta = 0.027$, $p = .106$. This implies that there is a (marginally) stronger relationship between similarity and dwell time for the KNN subjects than the BC subjects. We attempted to fit this regression with random slopes for subjects/films, as well, but this model did not converge, likely due to the size of the data (there are only so many observations at the subject-by-film level).

Another way to examine interesting trends is to let the eye-tracking data dictate the groupings. The second-block screen provided information falling into four distinct categories: (a) the current film, (b) the first-block films, (c) the first-block estimates, and (d) the first-block WTPs. We used each subject's average dwell proportions on each of these four categories to sort them into two clusters, using a standard k-means clustering classifier (with k = 2).

The average proportions for each of the two clusters and the strategy breakdown are in Table 1. Although not definitive, there is a divergence between the two strategies, in that most of the KNN subjects are sorted into the group that spends a greater proportion of time looking at the first-block films. This is a sensible finding, as KNN subjects should rely on the WTPs of the other films (in tandem with the similarity of these films) to update their current estimate. The BC subjects, on the other hand, fall roughly evenly into the two groups, suggesting perhaps that some of the BC subjects altered their strategy when provided with relevant onscreen information.

Indeed, when we resort the subjects according to their behavior in the second block, we see that 11 of the 21 subjects who were originally classified as BC were better explained by a KNN strategy. On the other hand, most (nine

Table 1
*K-Means Clustering*

| Group | Current film | Previous films | Estimates | WTPs | KNN | BC |
|---|---|---|---|---|---|---|
| 1 | 58% | 29% | 2% | 11% | 9 | 11 |
| 2 | 80% | 12% | 2% | 6% | 5 | 10 |

*Note.* WTP = willingness-to-pay; KNN = k-nearest neighbor; BC = beauty contest. The average proportion of trial-level dwell time spent looking at each of the four types of information on the screen, separated by cluster.

out of 14) subjects who were best fit by KNN in the first block were also best fit by the KNN model in the second block.

## Modeling in the Second Block

Before modeling in the second block, we calculated the optimal level of K in the first block (with a maximum of nine; Miller, 1956) for each of the KNN subjects; the optimal level of K is the one that minimizes *MSE* between the predicted and actual estimates. The crucial difference in the second block is the presence of explicit, onscreen feedback from the first block. Consequently, a primary question in this block is if and how subjects used the onscreen feedback. It makes sense that subjects would integrate their a priori estimate (based on their first-block strategy, i.e., KNN or BC) and the onscreen information using some sort of weighted average. The method in which subjects use the onscreen feedback, however, is not as obvious. Subjects might give more weight to feedback that they view earlier, later, or longer during the estimation process. Subjects might also weight the feedback according to the similarity between the gazed-at film and the target film.

To identify which of these methods subjects used, we tested 15 models at the individual level, each comprising weighted averages of the a priori first-block estimate ($Guess_{1st}$) and onscreen information-based estimates. The method by which the onscreen information is used differs between the models; each one includes a different combination of the three potential weighting mechanisms (order of gazes, duration of gazes, and similarity of gazed film to current film). The relative weight on the onscreen information ($w$) versus first-block guess ($1 - w$) was solved quadratically (Murty & Yu, 1988). More specifically, there is a closed-form solution to the weights ($w$ and $1 - w$) that most closely approximate the guesses made by subjects (i.e., those that minimize the squared errors). It is equivalent to fitting the following constrained regression:

$$Guess \sim \beta_0 + \beta_1 Guess_{1st} + \beta_2 Guess_{2nd}$$

where $\beta_0 = 0$, $\sum_{i=1}^{2} \beta_i = 1$, $Guess_{1\mathrm{st}}$ = first-block estimate, and $Guess_{2\mathrm{nd}}$ = gaze-informed second-block estimate. The first three models are the simplest, with only one attention-weighting mechanism per model. Each integrates one of (a) gaze primacy, (b) gaze recency, (c) gaze duration, or (d) similarity of the gazed-at film(s) into the estimate for that trial. In each model, $n$ is the total number of dwell in a trial, and *gaze* is an index for the current dwell. For each of the following 15 models, we normalized the onscreen information-based estimates to the appropriate range (i.e. 1–10). To do so, we divided the w-weighted summation by the sum of the non-WTP information. In the first two models, no normalization was necessary because the summation of the gaze order and gaze recency fractions are both 1. In the third model, for instance, we divided the w-weighted summation (the sum of the gaze duration-WTP products) by the sum of the trial-level gaze durations.

### (1) Simple gaze order

$$Guess_i = w\left(\sum_{gaze=1}^{n} \frac{2^{n-gaze}}{2^n - 1} * WTP_{gaze}\right) + (1 - w)(Guess_{1st})$$

This model assigns higher weight to earlier gazed-at movies.

### (2) Simple gaze recency

$$Guess_i = w\left(\sum_{gaze=1}^{n} \frac{2^{gaze-1}}{2^n - 1} * WTP_{gaze}\right) + (1 - w)(Guess_{1st})$$

This model assigns higher weight to later gazed-at movies.

### (3) Simple gaze duration

$$Guess_i = w\left(\sum_{gaze=1}^{n} Dur_{gaze} * WTP_{gaze}\right) + (1 - w)(Guess_{1st})$$

This model assigns higher weight to longer gazed-at movies.

**(4) Simple similarity**

$$Guess_i = w\left(\sum_{gaze=1}^{n} Sim_{gaze} * WTP_{gaze}\right) + (1 - w)(Guess_{1st})$$

This model assigns higher weight to more similar gazed-at movies.

The next six models combine two of the components from the simple models, with two attention weighting mechanisms per model. Each integrates two of (a) gaze primacy, (b) gaze recency, (c) gaze duration, or (d) similarity of the gazed film(s) into the estimate for that trial.

**(5) Gaze primacy and gaze recency**

$$Guess_i = w\left(\sum_{gaze=1}^{n} \frac{2^{n-gaze} + 2^{gaze-1}}{2(2^n - 1)} * WTP_{gaze}\right) + (1 - w)(Guess_{1st})$$

This model assigns higher weight to earlier and later gazed-at movies.

**(6) Gaze primacy and gaze duration**

$$Guess_i = w\left(\sum_{gaze=1}^{n} \frac{2^{n-gaze}}{2^n - 1} * Dur_{gaze} * WTP_{gaze}\right) + (1 - w)(Guess_{1st})$$

This model assigns higher weight to earlier and longer gazed-at movies.

**(7) Gaze recency and gaze duration**

$$Guess_i = w\left(\sum_{gaze=1}^{n} \frac{2^{gaze-1}}{2^n - 1} * Dur_{gaze} * WTP_{gaze}\right) + (1 - w)(Guess_{1st})$$

This model assigns higher weight to later and longer gazed-at movies.

**(8) Gaze primacy and similarity**

$$Guess_i = w\left(\sum_{gaze=1}^{n} \frac{2^{n-gaze}}{2^n - 1} * Sim_{gaze} * WTP_{gaze}\right) + (1 - w)(Guess_{1st})$$

This model assigns higher weight to earlier and more similar gazed-at movies.

**(9) Gaze recency and similarity**

$$Guess_i = w\left(\sum_{gaze=1}^{n} \frac{2^{gaze-1}}{2^n - 1} * Sim_{gaze} * WTP_{gaze}\right) + (1 - w)(Guess_{1st})$$

This model assigns higher weight to later and more similar gazed-at movies.

**(10) Gaze duration and similarity**

$$Guess_i = w\left(\sum_{gaze=1}^{n} Dur_{gaze} * Sim_{gaze} * WTP_{gaze}\right) + (1 - w)(Guess_{1st})$$

This model assigns higher weight to longer and more similar gazed-at movies.

The next set of models combine three out of the four potential attentional weighting mechanisms in the estimate for each trial.

**(11) Gaze primacy, gaze recency, and gaze duration**

$$Guess_i = w\left(\sum_{gaze=1}^{n} \frac{2^{n-gaze} + 2^{gaze-1}}{2(2^n - 1)} * Dur_{gaze} * WTP_{gaze}\right) + (1 - w)(Guess_{1st})$$

This model assigns higher weight to earlier, later, and longer gazed-at movies.

**(12) Gaze primacy, gaze recency, and similarity**

$$Guess_i = w\left(\sum_{gaze=1}^{n} \frac{2^{n-gaze} + 2^{gaze-1}}{2(2^n - 1)} * Sim_{gaze} * WTP_{gaze}\right) + (1 - w)(Guess_{1st})$$

This model assigns higher weight to earlier, later, and longer gazed-at movies.

**(13) Gaze primacy, gaze duration, and similarity**

$$Guess_i = w\left(\sum_{gaze=1}^{n} \frac{2^{n-gaze}}{2^n - 1} * Dur_{gaze} * Sim_{gaze} * WTP_{gaze}\right) + (1 - w)(Guess_{1st})$$

This model assigns higher weight to earlier, more similar, and longer gazed-at movies.

### (14) Gaze recency, gaze duration, and similarity

$$Guess_i = w\left(\sum_{gaze=1}^{n} \frac{2^{gaze-1}}{2^n - 1} * Dur_{gaze} * Sim_{gaze} * WTP_{gaze}\right) + (1 - w)(Guess_{1st})$$

This model assigns higher weight to later, more similar, and longer gazed-at movies.

The last model includes all four potential attentional weighting mechanisms in the estimate for each trial: gaze primacy, gaze recency, gaze duration, and similarity.

### (15) Gaze primacy, gaze recency, gaze duration, and similarity

$$Guess_i = w\left(\sum_{gaze=1}^{n} \frac{2^{n-gaze} + 2^{gaze-1}}{2(2^n - 1)} * Dur_{gaze} * Sim_{gaze} * WTP_{gaze}\right) + (1 - w)(Guess_{1st})$$

The *MSE*-minimizing model for each subject's guesses was identified, along with the best-fitting weight ($w$). These weights encompassed a wide range ([0.03, 0.82], 25th percentile = 0.25, 75th percentile = 0.62) and were quite variable ($M$ = 0.46, $SD$ = 0.23). A histogram of the distribution of weights can be found in Figure S4 in the online supplemental materials. The gaze recency (R), gaze duration (D), and similarity (S) variables seem to be more important than gaze primacy (P) in the estimation process. Out of all 35 subjects, only nine were best fit by a model that includes primacy. In contrast, 18 subjects' best-fitting models include gaze recency, 17 subjects' best-fitting models include gaze duration, and 17 subjects' best-fitting models include similarity (Table 2).

In addition to asking what subjects actually did in this study, we also found it important to ask what they *should* have done. Thus, we estimated the optimal weight for each subject in the same way that we estimated the actual weight ($w$). However, instead of comparing the predicted guesses to the subjects' actual guesses, we compared the predicted guesses to the actual WTPs. In doing so, we calculate how the subjects *should* have weighted the onscreen information to correctly guess the passive subject's WTP. Although subjects were far from being perfect estimators and although there was substantial variance in the weight subjects placed on the onscreen feedback, subjects weighted the onscreen information fairly appropriately. That is, there is a significant positive correlation between the $w$ they put on the onscreen feedback and the weight that would have optimized their guesses, given their pattern of gazes, $r(33) = 0.677$, $p = 10^{-6}$. The metrics for these models are provided in Table 3.

Interestingly, there is a clear split between the KNN and BC subjects in the actual, $t(24.189) = 7.598$, $p = 10^{-8}$, weights placed on the onscreen feedback. On average, BC subjects placed more weight on the onscreen information (i.e., had higher values of $w$; $M$ = 0.667, $SD$ = 0.130) than the KNN subjects ($M$ = 0.280, $SD$ = 0.158). Importantly, this is the pattern that we should expect to see because the optimal weights were higher, on average, for the BC subjects ($M$ = 0.563, $SD$ = 0.106) compared with the KNN subjects ($M$ = 0.209, $SD$ = 0.147), $t(21.755) = 7.753$, $p = 10^{-8}$.

Table 2
*Summary of Most-Used Model Parameters, Conditioned on Strategy*

| Model parameter | Number of subjects fit best by a model that includes that parameter: KNN | BC |
|---|---|---|
| Gaze primacy | 5 | 4 |
| Gaze recency | 9 | 9 |
| Gaze duration | 9 | 8 |
| Similarity | 8 | 9 |

*Note.* KNN = k-nearest neighbor; BC = beauty contest.

## Discussion

Overall, this study provides evidence that subjects can learn another person's preferences—at least, when given feedback about the accuracy of their past estimates. Individual differences exist, however, in the level of performance achieved and methods invoked by the subjects. In the more traditional learning setting (the first block), some subjects were best fit by a model that used feedback to adapt future estimates, whereas others' predictions fell more in line with a static strategy. This sorting extended to another context, yielding out-of-sample predictions about behavior and patterns of attention.

Table 3
*Metrics for Each Second-Block Model, Conditioned on Strategy*

| Model | Number of subjects fit best by this model | | Average actual weight | | Average optimal weight | | Average *MSE* | |
|---|---|---|---|---|---|---|---|---|
| | KNN | BC | KNN | BC | KNN | BC | KNN | BC |
| Gaze primacy | 1 | 0 | .17 | .57 | .15 | .50 | 2.77 | 4.74 |
| Simple gaze recency | 1 | 4 | .22 | .61 | .18 | .52 | 2.70 | 4.58 |
| Simple gaze duration | 2 | 4 | .26 | .66 | .20 | .57 | 2.68 | 4.43 |
| Simple similarity | 1 | 6 | .25 | .66 | .20 | .56 | 2.69 | 4.44 |
| Gaze primacy and gaze recency | 0 | 2 | .24 | .65 | .20 | .56 | 2.70 | 4.46 |
| Gaze primacy and gaze duration | 0 | 0 | .18 | .57 | .15 | .50 | 2.76 | 4.74 |
| Gaze recency and gaze duration | 0 | 1 | .22 | .60 | .18 | .51 | 2.70 | 4.59 |
| Gaze primacy and similarity | 0 | 1 | .17 | .57 | .14 | .49 | 2.76 | 4.77 |
| Gaze recency and similarity | 2 | 0 | .22 | .61 | .19 | .51 | 2.70 | 4.61 |
| Gaze duration and similarity | 1 | 1 | .25 | .64 | .18 | .55 | 2.69 | 4.48 |
| Gaze primacy, gaze recency, and gaze duration | 2 | 1 | .25 | .64 | .19 | .55 | 2.70 | 4.50 |
| Gaze primacy, gaze recency, and similarity | 0 | 0 | .23 | .64 | .19 | .55 | 2.70 | 4.51 |
| Gaze primacy, gaze duration, and similarity | 0 | 0 | .17 | .56 | .14 | .49 | 2.76 | 4.77 |
| Gaze recency, gaze duration, and similarity | 2 | 1 | .22 | .60 | .18 | .51 | 2.69 | 4.61 |
| Gaze primacy, gaze recency, gaze duration, and similarity | 2 | 0 | .23 | .62 | .18 | .54 | 2.70 | 4.56 |

*Note.* KNN = k-nearest neighbor; BC = beauty contest. The last two columns (average *MSE* for each model) show the deviations between the model and behavior; this suggests that the KNN model generally fits KNN subjects better than the BC model fits BC subjects.



Indeed, there were connections between the initial grouping and subjects' subsequent estimation processes.

Perhaps most interesting, however, is the finding that subjects weighted the onscreen feedback in the second block approximately optimally, given their pattern of gazes, the strategy into which they had been sorted in the first block, and the models tested herein. Despite the fact that some of the subjects were best fit initially by a static model (BC), both groups used the onscreen information—and the static subjects were found to have used it even more than their dynamic counterparts (KNN).

The higher weighting of onscreen information by BC subjects is especially interesting after considering the tendency of KNN subjects to spend a greater proportion of their estimation process looking at the external films. It is possible these models indicate that BC subjects put a greater weight on the onscreen information because their a priori estimates do not capture any of the passive subject's personal preferences, despite the fact that learning his preferences could be helpful. If, indeed, BC subjects put more weight on the on-screen information (on average, more than they did on their naïve, static strategy-based estimates), then they would have stood to improve their scores—at least to the level of the KNN subjects. This is one explanation for the performance difference between the strategies in the first block and the absence of such a difference in the second block.

Moreover, there is not a clear separation across groups regarding which variables were included in the best-fitting second-block models (Table 2). If the BC subjects had stuck to their static strategy, we might have expected far fewer of their best-fitting models to include the similarity variable. However, nine out of 21 BC subjects' (and eight out of 14 KNN subjects') best-fitting models include similarity. This provides some additional evidence that the BC subjects may have adapted their strategies during the second block of the experiment. It is interesting to note that the models with a single component (gaze primacy, gaze recency, gaze duration, or similarity) fit best for a majority (19/35) of subjects. We did not hypothesize this result a priori. However, it is possible that the interactions between components were not optimally specified. In other words, perhaps more subjects would have been best fit by one of the combination models if the model combined the components in a different way, for example, by weighting one more than the other(s) or allowing the interaction to be nonlinear. However, these adjustments would require additional parameters, which would increase the size of the model space very quickly. More research with a narrower focus would be required to fully address this possibility.

One interesting finding from the second-block model fits was that the primacy of information was rarely a useful parameter. This indicates that attention at the beginning of the estimation process is not as important as attention near the end of the estimation process or the overall amount of time spent on each piece of information. This is perhaps surprising, given that some choice models argue for the importance of early information in choice (e.g., models that incorporate lateral inhibition, as in Bogacz, Usher, Zhang, & McClelland, 2007; Usher & McClelland, 2001). However, it is not surprising that the gaze recency, gaze duration, and interfilm similarity variables are very common among best-fitting models. Gaze recency and duration were predicted to have an effect on the estimation process, in line with previous research regarding the role of attention in choice (Fiedler & Glöckner, 2012; Krajbich et al., 2010; Krajbich et al., 2012; Milosavljevic et al., 2012; Orquin & Mueller Loose, 2013; Pärnamets et al., 2015; Shimojo et al., 2003; Smith & Krajbich, 2018; Smith & Krajbich, 2019; Towal et al., 2013). Using similarity to inform the estimation process is seemingly the most "rational" of the three variables, and it is therefore reassuring to see that many subjects' processes were best fit by a model that includes similarity.

In addition, an important question is why subjects should use the onscreen information at all. One obvious reason is that memory is limited, so these films may actually provide useful information that would have otherwise not been remembered. Given the suboptimal BC strategy that many subjects seemed to use during the first block, it is possible that these subjects found it too difficult (or too effortful) to form a representation of the passive subject. In that case, adopting a more KNN-like strategy would only be feasible by utilizing the on-screen feedback, which BC subjects appeared to do in the second block to an even greater degree than the KNN subjects, who had ostensibly been able to form a representation of the passive subject during the first block. This discrepancy in the subjects' potential underlying strategies/tendencies could be a driving force in the large range of weights placed on the on-screen information.

However, it is important to recognize that the models suggested in the current study are by no means exhaustive of the potential strategies and processes used by subjects in a preference estimation task. For instance, the on-screen variables (gaze primacy, gaze recency, gaze duration, and similarity) are not necessarily the only variables involved in the estimation process. However, as a first attempt to understand the process underlying surrogate predictions, the range of strategies (with roots in learning and heuristics literature) and the incorporations of on-screen feedback (in accordance with research on attention and evidence accumulation) seem appropriately broad and grounded.

Moreover, it is important to recognize that the findings from this study are limited by the fact that there was only one passive subject. It is possible that the prediction processes for this passive subject would not generalize to other passive subjects. More specifically, because the passive subject's preferences were significantly positively correlated with the IMDb ratings, it is possible that the proportion of BC subjects is substantially higher than it would be otherwise. For instance, if a passive subject had no correlation between their WTP ratings and the IMDb ratings, then it is possible (or even probable) that fewer subjects would be sorted as BC. Thus, it is important to keep in mind that the distribution of strategies we observed may not be representative of other tasks. Our focus has been more on the link between strategies and process data, rather than the distribution of strategies per se.

An additional limitation in this article is the use of IMDb ratings as a proxy for the BC strategy. Though IMDb states that its ratings are not an exact average of all of the ratings for a given film (Ratings FAQ, n.d.), it is highly unlikely that the IMDb ratings are systematically biased from the average rating. In addition, to our knowledge, there is not a better alternative for film ratings. Rotten Tomatoes, for instance, gives an approval percentage, which is similar to but not the same as an average rating. Metacritic translates written reviews into a number, which might be prone to all sorts of issues (incorrect mapping, undue variability across genres, etc.). We considered translating the gross earnings from a film into a measure of liking, but that would be problematic as well because earnings are not necessarily correlated with quality or WTP. Therefore, we used IMDb ratings as an imperfect and useful proxy for how the "average" person feels about a film.

Ultimately, this project builds on past paradigms and findings (Janowski et al., 2013) and provides some novel groundwork for future research in the areas of social learning and surrogate decision-making. There is a growing body of literature investigating the differences between decisions made for oneself and those made for another. This research includes a variety of domains, including decisions over money (Albrecht, Volz, Sutter, Laibson, & von Cramon, 2011; Baldner, Longo, & Scott, 2015; Faro & Rottenstreich, 2006), relationships (Beisswanger, Stone, Hupp & Allgaier, 2003), physical safety (Stone, Choi, de Bruin, & Mandel, 2013), and medicine (Dore, Stone, & Buchanan, 2014; Fagerlin et al., 2001; Vig, Starks, Taylor, Hopley, & Fryer-Edwards, 2007). Generally, a common finding is that people choose differently for others than they would for themselves. Although the process underlying decisions for the self has been investigated in a variety of domains, the procedure invoked for other decisions has not received as much attention. Intuitively, when people make a decision for someone else, they try to take into account (to some extent, at least) the preferences of their recipient. Our work provides a framework for understanding the crucial step of how people learn these preferences.



## References

Aha, D. W., Kibler, D., & Albert, M. K. (1991). Instance-based learning algorithms. *Machine Learning, 6,* 37–66. http://dx.doi.org/10.1007/BF00153759

Aimone, J. A., Ball, S., & King-Casas, B. (2016). It's not what you see but how you see it: Using eye-tracking to study the risky decision-making process. *Journal of Neuroscience, Psychology, and Economics, 9,* 137–144. http://dx.doi.org/10.1037/npe0000061

Albrecht, K., Volz, K. G., Sutter, M., Laibson, D. I., & von Cramon, D. Y. (2011). What is for me is not for you: Brain correlates of intertemporal choice for self and other. *Social Cognitive and Affective Neuroscience, 6,* 218–225. http://dx.doi.org/10.1093/scan/nsq046

Ashby, N. J., Dickert, S., & Glöckner, A. (2012). Focusing on what you own: Biased information uptake due to ownership. *Judgment and Decision Making, 7,* 254–267.

Baldner, C., Longo, G. S., & Scott, M. D. (2015). The relative roles of drive and empathy in self- and other-focused financial decision making. *Personality and Individual Differences, 82,* 7–13. http://dx.doi.org/10.1016/j.paid.2015.01.025

Beisswanger, A. H., Stone, E. R., Hupp, J. M., & Allgaier, L. (2003). Risk taking in relationships: Differences in deciding for oneself versus for a friend. *Basic and Applied Social Psychology, 25,* 121–135. http://dx.doi.org/10.1207/S15324834BASP2502_3

Bogacz, R., Usher, M., Zhang, J., & McClelland, J. L. (2007). Extending a biologically inspired model of choice: Multi-alternatives, nonlinearity and value-based multidimensional choice. *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences, 362,* 1655–1670. http://dx.doi.org/10.1098/rstb.2007.2059

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10,* 433–436. http://dx.doi.org/10.1163/156856897X00357

Cavanagh, J. F., Wiecki, T. V., Kochar, A., & Frank, M. J. (2014). Eye tracking and pupillometry are indicators of dissociable latent decision processes. *Journal of Experimental Psychology: General, 143,* 1476.

Coricelli, G., & Nagel, R. (2009). Neural correlates of depth of strategic reasoning in medial prefrontal cortex. *Proceedings of the National Academy of Sciences of the United States of America, 106,* 9163–9168. http://dx.doi.org/10.1073/pnas.0807721106

Dore, R. A., Stone, E. R., & Buchanan, C. M. (2014). A social values analysis of parental decision making. *The Journal of Psychology: Interdisciplinary and Applied, 148,* 477–504. http://dx.doi.org/10.1080/00223980.2013.808603

Epley, N., & Gilovich, T. (2006). The anchoring-and-adjustment heuristic: Why the adjustments are insufficient. *Psychological Science, 17,* 311–318. http://dx.doi.org/10.1111/j.1467-9280.2006.01704.x

Fagerlin, A., Ditto, P. H., Danks, J. H., Houts, R. M., & Smucker, W. D. (2001). Projection in surrogate decisions about life-sustaining medical treatments.

*Health Psychology, 20,* 166–175. http://dx.doi.org/10.1037/0278-6133.20.3.166

Faro, D., & Rottenstreich, Y. (2006). Affect, empathy, and regressive mispredictions of others' preferences under risk. *Management Science, 52,* 529–541. http://dx.doi.org/10.1287/mnsc.1050.0490

Fiedler, S., & Glöckner, A. (2012). The dynamics of decision making in risky choice: An eye-tracking analysis. *Frontiers in Psychology, 3,* 335. http://dx.doi.org/10.3389/fpsyg.2012.00335

Fisher, G. (2017). An attentional drift diffusion model over binary-attribute choice. *Cognition, 168,* 34–45.

Gharib, A., Mier, D., Adolphs, R., & Shimojo, S. (2015). Eyetracking of social preference choices reveals normal but faster processing in autism. *Neuropsychologia, 72,* 70–79. http://dx.doi.org/10.1016/j.neuropsychologia.2015.04.027

Gill, D., & Prowse, V. (2017). *Using response times to measure strategic complexity and the value of thinking in games (SSRN Scholarly Paper No. ID 2902411)*. Rochester, NY: Social Science Research Network. Retrieved from https://papers.ssrn.com/abstract=2902411

Givi, J., & Galak, J. (2017). Sentimental value and gift giving: Givers' fears of getting it wrong prevents them from getting it right. *Journal of Consumer Psychology, 27,* 473–479. http://dx.doi.org/10.1016/j.jcps.2017.06.002

Hsee, C. K., & Weber, E. U. (1997). A fundamental prediction error: Self–others discrepancies in risk preference. *Journal of Experimental Psychology: General, 126,* 45–53. http://dx.doi.org/10.1037/0096-3445.126.1.45

Janowski, V., Camerer, C., & Rangel, A. (2013). Empathic choice involves vmPFC value signals that are modulated by social processing implemented in IPL. *Social Cognitive and Affective Neuroscience, 8,* 201–208. http://dx.doi.org/10.1093/scan/nsr086

Johnson, E. J., Camerer, C., Sen, S., & Rymon, T. (2002). Detecting failures of backward induction: Monitoring information search in sequential bargaining. *Journal of Economic Theory, 104,* 16–47. http://dx.doi.org/10.1006/jeth.2001.2850

Keynes, J. M. (1936). *General theory of employment, interest and money*. Cambridge, United Kingdom: Macmillan Cambridge University Press.

Kim, B. E., Seligman, D., & Kable, J. W. (2012). Preference reversals in decision making under risk are accompanied by changes in attention to different attributes. *Frontiers in Neuroscience, 6,* 109. http://dx.doi.org/10.3389/fnins.2012.00109

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., Broussard, C., and Associates. (2007). What's new in Psychtoolbox-3. *Perception, 36,* 1–16.

Knoepfle, D. T., Wang, J. T., & Camerer, C. F. (2009). Studying learning in games using eye-tracking. *Journal of the European Economic Association, 7,* 388–398. http://dx.doi.org/10.1162/JEEA.2009.7.2-3.388

Konovalov, A., & Krajbich, I. (2016). Gaze data reveal distinct choice processes underlying model-based and model-free reinforcement learning. *Nature Communications, 7,* 12438. http://dx.doi.org/10.1038/ncomms12438

Krajbich, I., Armel, C., & Rangel, A. (2010). Visual fixations and the computation and comparison of value in simple choice. *Nature Neuroscience, 13,* 1292–1298. http://dx.doi.org/10.1038/nn.2635

Krajbich, I., Hare, T., Bartling, B., Morishima, Y., & Fehr, E. (2015). A common mechanism underlying food choice and social decisions. *PLoS Computational Biology, 11,* e1004371. http://dx.doi.org/10.1371/journal.pcbi.1004371

Krajbich, I., Lu, D., Camerer, C., & Rangel, A. (2012). The attentional drift-diffusion model extends to simple purchasing decisions. *Frontiers in Psychology, 3,* 193. http://dx.doi.org/10.3389/fpsyg.2012.00193

Krajbich, I., & Smith, S. M. (2015). Modeling eye movements and response times in consumer choice. *Journal of Agricultural & Food Industrial Organization, 13,* 55–72. http://dx.doi.org/10.1515/jafio-2015-0016

Lindner, M. A., Eitel, A., Thoma, G.-B., Dalehefte, I. M., Ihme, J. M., & Köller, O. (2014). Tracking the decision-making process in multiple-choice assessment: Evidence from eye movements. *Applied Cognitive Psychology, 28,* 738–752. http://dx.doi.org/10.1002/acp.3060

Lohse, G. L., & Johnson, E. J. (1996). A comparison of two process tracing methods for choice tasks. *Organizational Behavior and Human Decision Processes, 68,* 28–43. http://dx.doi.org/10.1006/obhd.1996.0087

MathWorks. (2014). *MATLAB* (version 8.3). Natick, MA: The MathWorks Inc.

Miller, G. A. (1956). The magical number seven plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63,* 81–97. http://dx.doi.org/10.1037/h0043158

Milosavljevic, M., Navalpakkam, V., Koch, C., & Rangel, A. (2012). Relative visual saliency differences induce sizable bias in consumer choice. *Journal of Consumer Psychology, 22,* 67–74. Retrieved from http://dx.doi.org/10.1016/j.jcps.2011.10.002

Minogue, B. (1996). *Bioethics: A committee approach*. Retrieved from https://repository.library.georgetown.edu/handle/10822/889366

Moritz, B., Siemsen, E., & Kremer, M. (2014). Judgmental forecasting: Cognitive reflection and decision speed. *Production and Operations Manage-*

*ment, 23,* 1146–1160. http://dx.doi.org/10.1111/poms.12105

Murty, K. G., & Yu, F.-T. (1988). *Linear complementarity, linear and nonlinear programming* (Vol. 3). Berlin, Germany: Heldermann. Retrieved from http://mtzoumas.mysch.gr/mtz/docs_and_apps/Books/LCP_book.pdf

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General, 115,* 39–57. http://dx.doi.org/10.1037/0096-3445.115.1.39

Nosofsky, R. M. (2011). The generalized context model: An exemplar model of classification. In E. M. Pothos & A. J. Wills (Eds.), *Formal approaches in categorization* (pp. 18–39). Cambridge, United Kingdom: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511921322.002

Orquin, J. L., & Mueller Loose, S. (2013). Attention and choice: A review on eye movements in decision making. *Acta Psychologica, 144,* 190–206. http://dx.doi.org/10.1016/j.actpsy.2013.06.003

Pärnamets, P., Johansson, P., Hall, L., Balkenius, C., Spivey, M. J., & Richardson, D. C. (2015). Biasing moral decisions by exploiting the dynamics of eye gaze. *Proceedings of the National Academy of Sciences of the United States of America, 112,* 4170–4175. http://dx.doi.org/10.1073/pnas.1415250112

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10,* 437–442. http://dx.doi.org/10.1163/156856897X00366

Polonio, L., Di Guida, S., & Coricelli, G. (2014). Strategic sophistication and attention in games: An eye-tracking study. *Games and Economic Behavior, 94,* 80–96. http://dx.doi.org/10.1016/j.geb.2015.09.003

Ratings FAQ. (n.d.). *Internet Movie Database*. Retrieved from help.imdb.com/article/imdb/track-movies-tv/ratings-faq/G67Y87TFYYP6TWAV#ratings

Russo, J. E., & Leclerc, F. (1994). An eye-fixation analysis of choice processes for consumer nondurables. *Journal of Consumer Research, 21,* 274–290. http://dx.doi.org/10.1086/209397

Shimojo, S., Simion, C., Shimojo, E., & Scheier, C. (2003). Gaze bias both reflects and influences preference. *Nature Neuroscience, 6,* 1317–1322. http://dx.doi.org/10.1038/nn1150

Smith, S. M., & Krajbich, I. (2018). Attention and choice across domains. *Journal of Experimental Psychology: General, 147,* 1810–1826. http://dx.doi.org/10.1037/xge0000482

Smith, S. M., & Krajbich, I. (2019). Gaze amplifies value in decision making. *Psychological Science, 30,* 116–128. http://dx.doi.org/10.1177/0956797618810521

Stewart, N., Hermens, F., & Matthews, W. J. (2016). Eye movements in risky choice. *Journal of Behavioral Decision Making, 29*(2-3), 116–136.

Stone, E. R., Choi, Y., de Bruin, W. B., & Mandel, D. R. (2013). I can take the risk, but you should be safe: Self-other differences in situations involving physical safety. *Judgment and Decision Making, 8,* 250–267.

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement Learning: An Introduction*. Cambridge, MA: MIT Press.

Tamir, D. I., & Mitchell, J. P. (2010). Neural correlates of anchoring-and-adjustment during mentalizing. *Proceedings of the National Academy of Sciences of the United States of America, 107,* 10827–10832. http://dx.doi.org/10.1073/pnas.1003242107

Tavares, G., Perona, P., & Rangel, A. (2017). The attentional drift diffusion model of simple perceptual decision-making. *Frontiers in Neuroscience, 11,* 468. http://dx.doi.org/10.3389/fnins.2017.00468

Towal, R. B., Mormann, M., & Koch, C. (2013). Simultaneous modeling of visual saliency and value computation improves predictions of economic choice. *Proceedings of the National Academy of Sciences of the United States of America, 110,* E3858–E3867. http://dx.doi.org/10.1073/pnas.1304429110

Tversky, A., & Kahneman, D. (1975). Judgment under uncertainty: Heuristics and Biases. In D. Wendt & C. Vlek (Eds.), *Utility, probability, and human decision making* (pp. 141–162). Dordrecht, the Netherlands: Springer. http://dx.doi.org/10.1007/978-94-010-1834-0_8

Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review, 108,* 550–592. http://dx.doi.org/10.1037/0033-295X.108.3.550

Vaidya, A. R., & Fellows, L. K. (2015). Testing necessary regional frontal contributions to value assessment and fixation-based updating. *Nature Communications, 6,* 10120.

Vanderbilt, T. (2013). *The science behind the Netflix algorithms that decide what you'll watch next*. Retrieved from https://www.wired.com/2013/08/qq_netflix-algorithm/

Venkatraman, V., Payne, J. W., & Huettel, S. A. (2014). An overall probability of winning heuristic for complex risky decisions: Choice and eye fixation evidence. *Organizational Behavior and Human Decision Processes, 125,* 73–87. http://dx.doi.org/10.1016/j.obhdp.2014.06.003

Vig, E. K., Starks, H., Taylor, J. S., Hopley, E. K., & Fryer-Edwards, K. (2007). Surviving surrogate decision-making: What helps and hampers the experience of making medical decisions for others. *Journal of General Internal Medicine, 22,* 1274–1279. http://dx.doi.org/10.1007/s11606-007-0252-y

Wickelgren, W. A. (1977). Speed-accuracy tradeoff and information processing dynamics. *Acta Psychologica, 41,* 67–85. http://dx.doi.org/10.1016/0001-6918(77)90012-9

Received February 7, 2019
Revision received May 24, 2019
Accepted May 28, 2019 ■