Journal of Experimental Psychology: General

© 2017 American Psychological Association
0096-3445/17/$12.00 http://dx.doi.org/10.1037/xge0000382

Unrealistic Optimism in Advice Taking: A Computational Account

Yuan Chang Leong and Jamil Zaki
Stanford University

Expert advisors often make surprisingly inaccurate predictions about the future, yet people heed their suggestions nonetheless. Here we provide a novel, computational account of this unrealistic optimism in advice taking. Across 3 studies, participants observed as advisors predicted the performance of a stock. Advisors varied in their accuracy, performing reliably above, at, or below chance. Despite repeated feedback, participants exhibited inflated perceptions of advisors' accuracy, and reliably "bet" on advisors' predictions more than their performance warranted. Participants' decisions tightly tracked a computational model that makes 2 assumptions: (a) people hold optimistic initial expectations about advisors, and (b) people preferentially incorporate information that adheres to their expectations when learning about advisors. Consistent with model predictions, explicitly manipulating participants' initial expectations altered their optimism bias and subsequent advice-taking. With well-calibrated initial expectations, participants no longer exhibited an optimism bias. We then explored crowdsourced ratings as a strategy to curb unrealistic optimism in advisors. Star ratings for each advisor were collected from an initial group of participants, which were then shown to a second group of participants. Instead of calibrating expectations, these ratings propagated and exaggerated the unrealistic optimism. Our results provide a computational account of the cognitive processes underlying inflated perceptions of expertise, and explore the boundary conditions under which they occur. We discuss the adaptive value of this optimism bias, and how our account can be extended to explain unrealistic optimism in other domains.

Keywords: advice-taking, computational modeling, confirmation bias, optimism bias, social learning

Supplemental materials: http://dx.doi.org/10.1037/xge0000382.supp

In a series of forecasting tournaments conducted between 1984 and 2003, 284 experts including well-known political analysts, economists, and journalists were recruited to predict the likelihood that different political events would occur within a specified timeframe (e.g., Will the United States go to war in the Persian Gulf in the next 10 years?). These tournaments resulted in 28,000 specific, testable predictions, which rarely performed better than chance (Tetlock, 2005). Despite the poor performance, individuals who made these predictions were described as experts by the media, and their predictions heeded by policymakers. Investors likewise put their faith in financial "gurus" who are at chance at predicting the market (Shefrin, 2000), and consumers adopt questionable health practices recommended by medical talk shows (Korownyk et al., 2014). Even in laboratory experiments, human participants often follow misleading advice, leading to poor decisions (Biele, Rieskamp, & Gonzalez, 2009; Doll, Jacobs, Sanfey, & Frank, 2009; Staudinger & Büchel, 2013).

In all these cases, people trust others' opinions more than they should, a phenomenon we term *optimism bias in advice taking*. Why do decision-makers[1] exhibit this bias? Here we explore two potential sources of undue optimism in social settings—biased initial expectations and confirmation bias. In doing so, we focus on decision-makers' beliefs about an advisor's ability to make accurate forecasts of future events.

Biased Initial Expectations

Decision-makers often have to quickly judge an advisor's expertise before deciding whether to act on her advice. As with other first impressions, people often rely on superficial cues to assess expertise (Bonaccio & Dalal, 2006; Hovland, Janis, & Kelley, 1953). For instance, decision-makers privilege credentials (e.g., passing a certification examination), experience in a related domain, and prestige (e.g., working for a renowned firm) when deciding how accurate advisors are likely to be (Berlo, Lemert, & Mertz, 1969; Birnbaum & Stegner, 1979). Expert advisors are often portrayed to have privileged knowledge of their domain of

Yuan Chang Leong and Jamil Zaki, Department of Psychology, Stanford University.

Data for all three experiments, example stimuli, and custom code for computational models are available on GitHub: https://github.com/ycleong/AdviceTaking.

These results were previously presented in a poster at the 2016 Society for Personality and Social Psychology Annual Meeting in San Diego, CA, and as a symposium talk at the 2017 Society for Personality and Social Psychology Annual Meeting in San Antonio, TX.

Correspondence concerning this article should be addressed to Yuan Chang Leong, Department of Psychology, Stanford University, 450 Serra Mall, Jordan Hall, Stanford University, Stanford, CA 94305-2130. E-mail: ycleong@stanford.edu

[1] For clarity of prose, we will be using the term *decision-maker* to refer to the individual receiving advice. Several other equivalent terms have been used in the related literature, including *judge, client, advice-seeker* and *principal*. We will refer to the individual providing advice as the *advisor*.

expertise. In these cases, decision-makers are *predisposed* to believe in advisors' expertise, even before these advisors make any specific predictions. These initial expectations can bias decision-makers toward taking even inaccurate advice (cf. Tetlock, 2005).

Biased Learning

Even in the face of initial expectations, decision-makers adjust their expectations about advisors over time, for instance increasingly relying on advisors who have proved accurate in the past (Yaniv & Kleinberger, 2000). However, decision-makers' learning can be systematically *biased*. In particular, the desire to maintain consistency in their beliefs leads individuals toward *confirmation bias*: overweighting information that confirms their expectations, while discounting disconfirming information (Kunda, 1990; Nickerson, 1998; Oswald & Grosjean, 2004).

In the context of advice taking, confirmation bias could compound the effects of initial optimism. Consider an advisor who makes an erroneous prediction. The advisor might be inept, or simply unlucky. How would a decision maker interpret the advisor's error? One possibility is that decision-makers who already believe the advisor to be accurate would more readily attribute his errors to chance, while attributing accurate predictions to his ability.

A Computational Approach

To tease apart the influence of initial expectations versus confirmation bias, one must dynamically track decision-makers' beliefs about an advisor's expertise. One approach for doing so entails building computational models that make trial-by-trial estimates about decision-makers' beliefs, and testing those models against experimental data. Such models have successfully uncovered mechanisms through which people form beliefs about others' emotional states (Ong, Zaki, & Goodman, 2015), goals (Baker, Saxe, & Tenenbaum, 2009), and intentions (Diaconescu et al., 2014). Here, we likewise use computational models to (a) estimate decision-makers' initial expectations, (b) dynamically track how these expectations shift over time, and (c) evaluate the role of biased expectations and learning in overly optimistic advice taking.

The Present Study

We adapted an experimental task that mimicked real-world financial advice-taking (Boorman, O'Doherty, Adolphs, & Rangel, 2013). In the *Stock Prediction* phase of the task, participants predicted whether a fictitious stock would increase or decrease in price across successive time periods. In the *Advisor Evaluation* phase, participants observed financial advisors making predictions about a different stock, and made bets on whether these predictions would be correct or incorrect. In the *Joint Prediction* phase, they received recommendations from the advisors whom they had observed in the *Advisor Evaluation* phase, and predicted the performance of a third stock after receiving that advice.

In Experiment 1, we evaluated participants' perceptions of advisor expertise, and their utilization of advice in making stock predictions. We hypothesized that participants would exhibit an optimism bias, relying on advice more than they should. We fit computational models to determine whether these biases were attributable to optimistic initial expectations, confirmation bias or a combination of both factors. In Experiment 2, we explicitly manipulated participants' initial expectations by providing false information about the advisors, and assessed if and how these expectations affected participants' decisions. Finally, in Experiment 3, we investigated whether the optimism bias "spreads" across generations of decision-makers. One group of participants performed the financial advice-taking task and rated each advisor from one to five stars based on their perception of the advisor's expertise. The average ratings of each advisor were then passed to a second group of participants prior to them performing the task.

Together, the current work provides a computational account of an optimism bias in advice-taking, and explores the conditions under which unrealistic optimism occurs.

Experiment 1

In Experiment 1, participants learned about the expertise of advisors and later made decisions about utilizing each advisor's advice when making stock predictions. This task allowed us to examine whether participants overestimated advisors' expertise, and whether they utilized advice more often than they should. We also formalized a computational model of how participants learned about advisors' expertise, which we then used to separately examine the role of initial expectations and confirmation bias in contributing to optimistic advice taking.

Method

Participants. Twenty-seven participants were recruited from the Stanford community (18 male, 9 female, ages 19–43, mean age = 24.2). All participants provided written, informed consent prior to the start of the study. All experimental procedures were approved by the Stanford Institutional Review Board. Participants were paid up to $16 depending on their performance on the task. We discarded data from one participant who missed more than 10% of the trials, yielding a final sample of 26 participants.

Stimuli. Face stimuli of White male faces posing calm expressions with mouth closed and eyes gazing straight ahead were taken from the IASLab Face Set,[2] and used as photos for advisors. Stimuli were presented using MATLAB software (MathWorks) and the Psychophysics Toolbox (Brainard, 1997). Example stimuli can be viewed on our GitHub repository (https://github.com/ycleong/AdviceTaking#example-face-stimuli).

Experimental task and design. The Financial Advice and Choice task (FAC task, Figure 1A, adapted from Boorman et al. (2013)), consists of three phases. In the *Stock Prediction* phase, participants were asked to predict the price fluctuation of a fictitious stock based on the stock's past history. In each time period, participants had to predict whether the price of the stock would go up or down in the next time period. Participants had 3 seconds to respond before the trial timed out. Once they made their prediction, they were shown the actual performance of the stock (1.2

[2] Development of the Interdisciplinary Affective Science Laboratory (IASLab) Face Set was supported by the National Institutes of Health Director's Pioneer Award (DP1OD003312) to Lisa Feldman Barrett. More information is available online at www.affective-science.org.

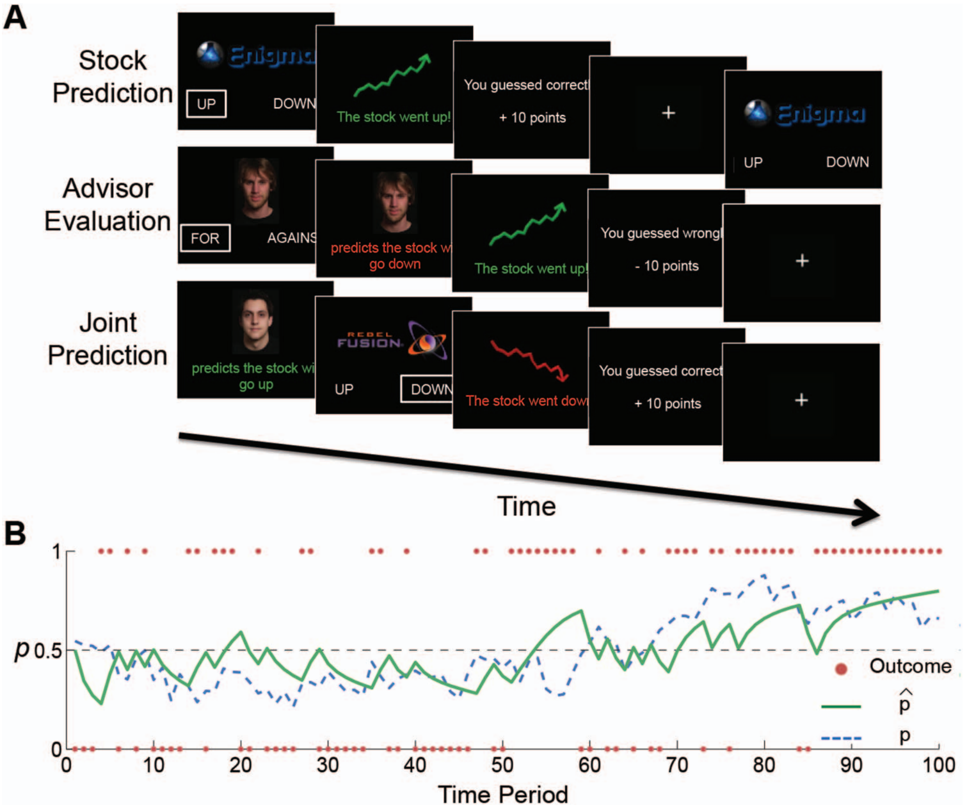

Figure 1. Task Design. (A) Financial advice choice (FAC) task. *Stock Prediction Phase*. Participants were tasked to predict the price fluctuation of a stock in consecutive time periods. *Advisor Evaluation Phase*. Participants observed financial advisors perform the same stock prediction task. In each time period, participants had to bet on whether the advisor would make a correct prediction. Participants observed 3 different advisors, with accuracies of 75%, 50% and 25% respectively. *Joint Prediction Phase*. Participants performed the stock prediction task on a third stock. Prior to making their predictions, participants received a recommendation from an advisor they observed in the Advisor Evaluation phase. (B) Stock Behavior. On each time period, stock performance (red dots, 1 = up, 0 = down) is determined probabilistically. The probability that the price of the stock will increase drifts from time period to time period (p, blue dotted line). The Bayesian Learning model is used to obtain the best estimate of p given the past outcomes of the stock ($\hat{p}$, green solid line). See the online article for the color version of this figure.

seconds) followed by feedback as to whether their prediction was right or wrong (1.2 seconds). After a 1-s interval, they moved onto the next trial. Participants began with an endowment of 800 points, and earned 10 points for each correct prediction, and lost 10 points for each incorrect prediction. At the end of the experiment, the points were converted to money (10 points = $0.10) to determine participants' payment.

Participants made predictions for the same stock over 100 consecutive time periods, with a self-paced break at time period 50. We manipulated the performance of the stock such that the probability that the stock would increase in value, p, "drifted" across time (Figure 1B). More specifically, p was initialized to a random value between 0 to 1. At each subsequent time period t, p was drawn randomly from a beta distribution with mean set to the p of the previous time period ($t - 1$) and a standard deviation of 0.07. Effectively, this means that when the price of a stock increases, it tends to keep increasing, and when the price of a stock decreases, it tends to keep decreasing. Previous work using similar setups has shown that participants are able to pick up on these trends to make accurate predictions (Behrens, Hunt, Woolrich, & Rushworth, 2008; Behrens, Woolrich, Walton, & Rushworth, 2007; Boorman et al., 2013). The *Stock Prediction* phase provided participants with an opportunity to familiarize themselves with the dynamics of the stock trend before the subsequent phases of the task. We examined participants' predictions in the *Stock Prediction* phase to evaluate whether participants relied on the stock trend to make accurate predictions (see *Analysis of Stock Prediction and Joint Prediction phases*).

In the *Advisor Evaluation* phase, participants observed as financial advisors predicted the price fluctuations of a different stock. The performance of the stock was manipulated using the same procedure. In each time period, participants were presented with a photograph of one of three male Caucasian faces, each representing a financial advisor. Participants were then asked to bet *for* the advisor—guessing that his next prediction would be accurate—or to bet *against* him—guessing that it would be inaccurate. Participants had 3 seconds to make their bet. They were then shown the advisor's prediction (1.2 seconds) followed by the actual performance of the stock (1.2 seconds). Participants were then given feedback about their bets—they earned 10 points for correct bets and lost 10 points for incorrect bets (1.2 seconds). After a 1-s interval, they moved onto the next time period.

Each participant encountered three advisors with different levels of expertise. One advisor made accurate predictions in 75% of the time periods (75% Advisor). Another advisor was at chance at making predictions (50% Advisor), and a third advisor was accurate in 25% of the time periods (25% advisor). For each participant, the photograph associated with each accuracy level was randomly selected from a subset of faces from the IASLab Face Set. Participants were not given any information about the expertise of the advisors, and thus had to learn about the advisors over time. Participants made bets for 108 time periods (36 time periods with each advisor). As participants made their bets *before* seeing the advisor's prediction, the bets are not influenced by beliefs about the stock trend. Instead, they provide us with a proxy measure of participants' beliefs about each advisor's accuracy and how these beliefs evolved over time. The order of advisors was pseudorandomized such that the transition probability between any pair of advisors was equated (e.g., a 75% advisor was equally likely to be followed by a 50% advisor as a 25% advisor).

In the *Joint Prediction* phase of the task, participants predicted the performance of a third fictitious stock. At the start of each trial, participants received a recommendation from one of the advisors they had encountered in the *Advisor Evaluation* phase. A photo of the advisor was presented, along with a prediction of whether the price of the stock would increase or decrease (1.2 seconds). After seeing the advisor's recommendation, participants were given 3 seconds to predict whether the stock's price would increase or decrease. Participants were then told whether their prediction was correct (1.2 seconds), and if they earned or lost points (1.2 seconds). Notably, in this phase, participants could make their predictions based on two sources of information: the advisor's prediction, and the stock's performance over recent time periods. Participants made predictions about the stock for 216 time periods (72 with each advisor). The order of advisors was again pseudorandomized to equate the transition probability between advisors.

At the end of the experiment, participants were asked how accurate they thought each advisor was at making predictions about the stock. Participants were asked to enter a percentage from 0–100%.

Learning about advisors' expertise. We operationalized participants' beliefs about an advisor's expertise through their bets during the Advisor Evaluation phase. Specifically, we calculated the proportion of trials on which participants bet *for* each advisor's prediction, and used robust Bayesian estimation to examine if the proportions were different from chance (see *Robust Bayesian estimation* for details).

Computational modeling. The proportion of bets alone, however, ignores the temporal dynamics of participants' beliefs about advisors as they learn more about their performance. To better capture these dynamics, we fit participants' bets to a set of computational models. Importantly, we fit two models that differed in their assumptions about how decision-makers learn about the advisors based on feedback. Our *Bayesian Learning model* assumes that participants learned in a statistically optimal fashion, whereas our *Confirmation Bias model* instead assumed that participants preferentially learn from evidence that accords with their previous beliefs. Comparing the fit of each model to participants' data allowed us to assess whether participants indeed exhibit biases in learning about advisors. Further, each model produced an estimate for participants' *priors*, or initial expectations, about an advisor's expertise. These priors offer a quantitative assessment of the extent to which observers exhibit unreasonable optimism about advisors before encountering evidence about their performance.

Bayesian Learning model. The Bayesian Learning model assumes that participants update their beliefs about each advisor's expertise in a statistically optimal manner, in accordance with Bayes rule:

$$\underset{\textit{posterior}}{\mathrm{P(expertise}\,|\,\mathrm{outcome)}} \propto \underset{\textit{prior}}{\mathrm{P(expertise)}} \times \underset{\textit{likelihood of observed data}}{\mathrm{P(outcome}\,|\,\mathrm{expertise)}}$$

This class of models has been found to accurately describe how individuals learn the probability of correct advice from an advisor (Behrens et al., 2008), reward probabilities in the environment (Behrens et al., 2007) and the association between visual cues and task contexts (Waskom, Frank, & Wagner, 2017).

On each trial, participants have an existing, or *prior*, belief about an advisor's expertise that can be represented as a probability distribution between 0 (nonexpert) and 1 (expert; Figure 2A). Expertise is linearly related to the likelihood that the advisor will provide accurate advice (Figure 2B). When participants observe the outcome of an advisor's predictions, they consider the *likelihood* of that outcome given the advisor's expertise. For example, an inaccurate prediction would be more likely with a nonexpert advisor than an expert advisor. Bayes rule provides a mathematically precise method for computing the updated, or *posterior*, belief—by taking the normalized product of the prior distribution and the likelihood function (Figure 2D, blue line). The posterior belief on a given current trial then becomes the prior for the subsequent trial, thus capturing participants' evolving beliefs about advisor expertise.

Our Bayesian model assumes that participants weigh recent experiences more than distant ones. This is implemented via the inclusion of a parameter v that estimates the *volatility* in an advisor's accuracy, and determines the rate at which the estimates of an advisor's expertise change from trial-to-trial (Behrens et al., 2007; Waskom et al., 2017). When an advisor is highly erratic (i.e., quick alternations between being accurate and inaccurate), v is high, and recent experiences with the advisor would be weighted more heavily, resulting in estimates of advisor expertise changing quickly from trial-to-trial. Alternatively, when an advisor is consistent, v is low, and estimates of advisor expertise change less with each new observation. As such, v can be thought as an "optimal learning rate" estimated using Bayesian inference.

Formally, the model can be written as:

$$p(a_t, v_t \,|\, y_{1:t}) \propto p(y_t \,|\, a_t) \int p(a_{t-1}, v_{t-1}) | \, y_{1:t-1} p(a_t \,|\, a_{t-1}, v_t) \, da_{t-1}$$

where α denotes advisor expertise, y denotes an observed outcome (accurate prediction: $y = 1$, inaccurate prediction: $y = 0$) and $p(a_{t,} v_t | y_{1:t})$ denotes the posterior probability of advisor expertise and volatility, after having observed outcomes from time period 1 to time period t. At each time period, the posterior estimate of the advisor's expertise was approximated using grid sampling, and the continuous distributions were discretized to allow for numerical integration We can then marginalize over v to obtain a posterior

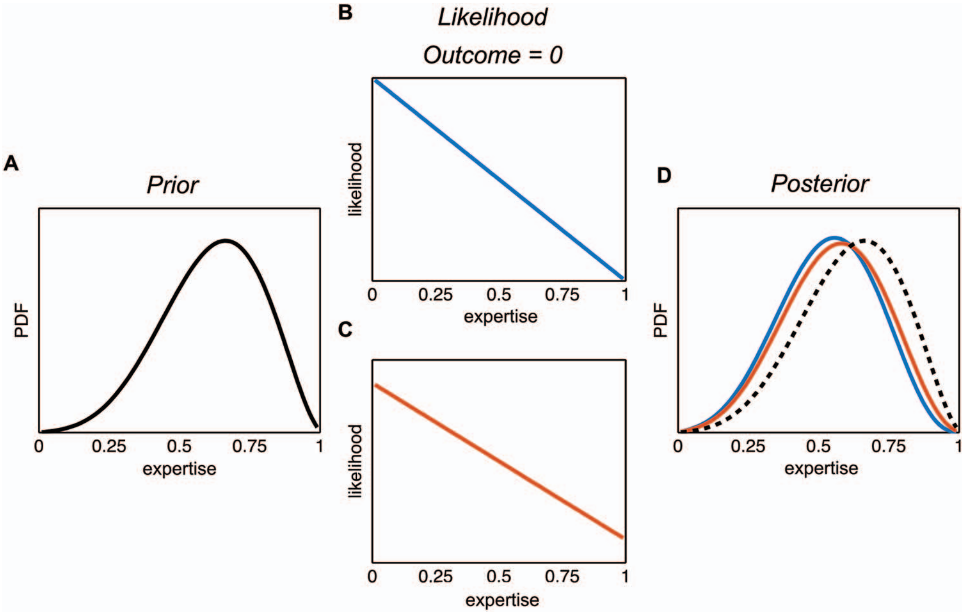

Figure 2. Comparison between the Bayesian Learning model and the Confirmation Bias model. (A) Probability distribution representing participants' *prior* belief about the advisor's expertise. The distribution has greater mass on the right, indicating an optimistic expectation. (B) *Likelihood* function for the Bayesian learning model when the advisor makes an inaccurate prediction. An advisor with low expertise is more likely to provide an inaccurate prediction than an advisor with high expertise. (C) *Likelihood* function for the Confirmation Bias model when an advisor, whom participants expect to be accurate, makes an inaccurate prediction. There is decreased likelihood for an inaccurate prediction at lower levels of expertise, implying that the inaccurate prediction is less diagnostic of low expertise. (D) Probability distributions representing the *posterior* belief about the advisor's expertise for the Bayesian Learning model (blue/dark grey) and the Confirmation Bias model (red/light grey). The prior belief in 2A is plotted again (black dotted line) for comparison. The decrement in participants' estimate of advisor's expertise is smaller for the Confirmation Bias model than the Bayesian Learning model after observing inaccurate advice. See Fig. S2 for a trial-by-trial comparison of the two models over 10 consecutive time periods. See the online article for the color version of this figure.

distribution for advisor expertise (e.g., Figure 2D), and compute the mean of that distribution as the best estimate of α_t. The full algebraic formulation of the model is provided in the supplemental materials.

Our model differs from previous work in its treatment of how participants' initial belief about advisors is specified. Whereas previous work assumes a uniform prior, reflecting no prior knowledge about the advisor, we hypothesized that participants begin the experiment with preconceived beliefs about advisor expertise. Hence, instead of explicitly specifying an initial belief distribution, we fit the model to participants' data to find the initial belief distribution that would give rise to the best fit to the data. Specifically, we assumed that participants' initial belief over α_1 can be described by a Beta distribution:

$$\alpha_1 \sim Beta(\alpha, \beta)$$

The shape (i.e., skew, mean, variance) of a Beta distribution depends on two parameters, α and β. The mean estimate of the distribution (i.e., the *expected* value of α_1, denoted here as $\hat{\alpha}_1$) can be calculated as:

$$\hat{\alpha}_1 = \frac{\alpha}{\alpha + \beta}$$

By varying α and β, we can define a Beta distribution with more mass on the right (i.e., $\hat{\alpha}_1 > 0.5$, "optimistic" beliefs) or a Beta distribution with more mass on the left (i.e., $\hat{\alpha}_1 < 0.5$, "pessimistic" belief). When $\alpha > \beta$, the distribution has greater mass on the right, indicating an optimistic belief in the advisor's expertise. As an example, Figure 2A shows an optimistic belief distribution defined by $Beta(5{,}3)$, with $\hat{\alpha}_1 =$ of 0.625. Conversely, if $\beta > \alpha$, the distribution has greater mass on the left, indicating a pessimistic initial belief about the advisor's expertise. We fit the model to data in the *Advisor Evaluation* phase to find the values of α and β that provided the best fit to each participant's responses (see *Model fitting and comparison* section below). We then take the Beta distribution parameterized with the best-fit values of α and β to be our estimate of that participant's prior beliefs about advisors' expertise.

Confirmation Bias model. The Confirmation Bias model modifies the Bayesian Learning Model to assume that participants "explain away," or underweight, new information contradicting their expectations, such as an inaccurate prediction from an advisor whom participants expect to be accurate. Specifically, the model computes a new likelihood function, $p_b(\text{outcome} \mid \text{expertise})$, that is the weighted combination of the likelihood of the observed outcome and that of the expected outcome (Figure 2C, see also supplemental methods):

if $\hat{\alpha}_i > 0.5$,

$$p_{bi}(\text{outcome} = 0 \mid \text{expertise}) \propto$$

$$b_i \times \text{p(outcome} = 1 \mid \text{expertise)} + (1 - b_i) \times \text{p(outcome} = 0 \mid \text{expertise)}$$

or

if $\hat{\alpha}_i < 0.5$,

$$p_{bi}(\text{outcome} = 1 \mid \text{expertise}) \propto$$

$$b_i \times \text{p(outcome} = 0 \mid \text{expertise)} + (1 - b_i) \times \text{p(outcome} = 1 \mid \text{expertise)}$$

where p_{bi}(outcome | expertise) denotes the likelihood function used by the Confirmation Bias model on trial i, $\hat{\alpha}_i$ denotes participants' mean estimate of the advisor's accuracy on trial i, and b_i is a bias term that weights the influence of expectations, and scales linearly with participants' current estimates about the advisor:

$$b_t = |\hat{\alpha}_t - 0.5|$$

such that the magnitude of bias is greater when participants expect advisors to be highly accurate (high $\hat{\alpha}_i$) or highly inaccurate (low $\hat{\alpha}_i$). As such, depending on participants' current expectations about an advisor's expertise, the model assigns different likelihoods to receiving accurate or inaccurate advice from an advisor (Fig. S1). When participants have *optimistic expectations* (i.e., $\hat{\alpha}_i > 0.5$), the likelihood of receiving inaccurate advice from an inaccurate advisor is decreased, and the likelihood of receiving inaccurate advice from an accurate advisor is increased (Figure 2C). This implies that accurate advice is weaker evidence, or less diagnostic, of low expertise. In contrast, if participants had *pessimistic expectations* (i.e., $\hat{\alpha}_i < 0.5$), the likelihood of receiving accurate advice from an accurate advisor is decreased, and the likelihood of receiving accurate advice from an inaccurate advisor is increased, implying that accurate advice is less diagnostic of high expertise.

Model-fitting and model comparison. To find the prior distribution that best describes participants' initial belief about advisors' expertise, we fit both the Bayesian Learning and Confirmation Bias models to participants' bets in the Advisor Evaluation phase to find the best-fit values of α and β for each participant. As participants make their bets *before* seeing the advisor's stock prediction, knowledge of the stock trend does not help participants in making their bets. Instead, participants ought to bet for or against an advisor's prediction based on their beliefs about the advisor's expertise. We assumed that the relationship between participants' estimate of an advisor's expertise and their bets on the advisor's prediction is described by a logistic function:

$$p(bet_i = FOR) = \frac{1}{1 + e^{-\tau(\hat{\alpha}_i - 0.5)}}$$

where $\hat{\alpha}_i$ denotes participants' mean estimate of the advisor's expertise on trial i, and τ is a subject-specific free parameter that determines the gain of the logistic function. τ allows the choice rule to interpolate between a maximization rule (i.e., always betting for the advisor's prediction when $\hat{\alpha}_i > 0.5$, and always betting against the advisor's prediction when $\hat{\alpha}_i < 0.5$), a "soft" maximization rule that assumes participants are more likely to bet for the advisor when $\hat{\alpha}_i$ is high, and a random choice rule (Fig. S3). In particular, for large values of τ, the choice rule approaches maximization, and for small values of τ, the choice rule approaches random choice. For moderate values of τ, the choice rule weights the probability of betting for the advisor's prediction by the estimate of the advisor's accuracy. As τ is fit to each individual participant separately, the logistic function allowed us to capture a spectrum of choice strategies across participants.

Accordingly, we can fit the models to each participant's data to find the best-fit values of α, β and τ. Because α, β, and τ can theoretically take on any values from 0 to positive infinity, we imposed regularizing priors (X ~ Gamma(2,3)) to facilitate realistic values during model fitting (Daw, 2011). We then fit the model to find the values of α, β and τ that maximize the posterior probability of the data given the model. These best-fit values are referred to as the *maximum a posteriori (MAP) parameter estimates.*

We evaluated model fit based on the average likelihood per trial, corrected for the number of free parameters in the model. The corrected average likelihood per trial is derived from the Akaike Information Criterion (AIC; Akaike, 1974).

$$AIC_c = -2\log lik + 2k + \frac{2k(k+1)}{N-k-1}$$

where AIC_c refers to the finite sample size corrected version of AIC recommended for small data sets (Burnham & Anderson, 2004; Hurvich & Tsai, 1989), *lik* is the maximum likelihood of the data given the model, k is the number of free parameters, and N is the number of data points. The second term serves as a penalty term that scales with the number of free parameters, and the third term adds a correction for finite-sample biases. From the AIC_c, we computed an unbiased estimate of the expected log likelihood of out-of-sample data given the model (Akaike, 1978; Gelman, Hwang, & Vehtari, 2014):

$$\log L = -\frac{1}{2}AIC_c$$

which was then divided by the number of trials, and exponentiated to obtain the corrected average likelihood per trial given the model:

$$\text{average likelihood} = \exp\left(\frac{\log L}{N}\right)$$

The corrected average likelihood per trial denotes the average likelihood of predicting a new data point given the model. It varies between 0 and 1, with 0.5 indicating chance likelihood and 1 indicating perfect correspondence. Notably, as the average likelihood measure was derived from the expected log likelihood of out-of-sample data, it is a measure of model fit that has been corrected for the number of free parameters and can be used to compare between models. We computed the average likelihood per trial separately for each participant. To compare between the model fits of the Confirmation Bias and Bayesian Learning models, we applied robust Bayesian estimation to assess if the within-participant differences in corrected average likelihood per trial were credibly different from 0 (see *Robust Bayesian Estimation* below). In examining within-participant difference in model fits rather than aggregate model performance, we treat model fit as a random effect, with the implicit assumption that the best-fitting model might differ between participants.

Model simulations. We performed a simulation study to isolate the pattern of results we would observe if participants' behavior were perfectly described by each model. We simulated the Confirmation Bias and Bayesian Learning models performing the Advisor Evaluation phase of the task with the best-fit parameters, and examined whether simulated model behavior replicated the pattern of results observed in the data. For each participant, we generated 36 predictions each from a 75% accurate advisor, a 50% accurate advisor and a 25% accurate advisor. We then simulated the bets that the models would make on each time period, given that participant's best-fit values of α, β, and τ. We repeated the procedure 500 times, and computed the average number of time periods on which the models bet for each advisor.

In addition, to better visualize the differences between the two models, we simulated the two models learning about an advisor with chance accuracy over the course of 10 time periods. On each trial, we plot the posterior distribution over the advisor's expertise, and observed how the behavior of the two models diverged over time (Fig. S2, see *Trial-by-trial comparison of Bayesian Learning model and Confirmation Bias model* in supplemental materials). Finally, we ran a parameter recovery study to examine whether the two models are identifiable. We fit the models to simulated data with known parameter values and showed that we were able to accurately recover the true values of α_1 and τ (Fig. S3, see *Parameter recovery study* in supplemental materials).

Analysis of stock prediction and joint prediction phases. Participants made predictions about the price fluctuations of a stock in both the *Stock Prediction* and *Joint Prediction* phases. In the *Stock Prediction* phase of the task, participants made their predictions without recommendations from advisors, and had to rely solely on the stock trend. As one measure of participants' performance, we computed the percentage of trials on which they correctly predicted the stock. As a second measure of participants' performance, we quantified the percentage of trials on which their bets were consistent with the stock trend—betting for the stock when it was likely to rise in value ($p > .5$) and against it when it was likely to fall ($p < .5$). To do so, we estimated p by applying the Bayesian Learning model to the history of the stock outcomes until that trial. We then used a mixed effects logistic regression model to predict participants' stock predictions from the estimated stock trend. 95% confidence intervals (CI) for the regression coefficients were computed using a parametric bootstrap analysis with 500 iterations, and are reported in square brackets next to the estimated coefficients.

In the *Joint Prediction* phase, participants again predicted the price fluctuation of another fictitious stock. However, unlike the *Stock Prediction* phase, participants could now rely on two distinct pieces of information when making their predictions: the stock trend based on the previous outcomes of the stock, and a recommendation from one of the advisor they observed in the *Advisor Evaluation* phase. To evaluate how much participants weighed each source of information, we ran a mixed effects logistic regression model that predicted participants' stock predictions from the stock trend (as estimated by the Bayesian Learning model) and the recommendation from each advisor. We z-scored each variable prior to entering it into the regression, and obtained the standardized beta coefficients for p and the recommendations from each advisor. This regression allowed us separately to test (a) whether participants weighted the stock trend more or less than the advice they received, (b) whether participants weight advice based on the relative accuracies of each advisor, and (c) whether participants overweight the advice of an advisor with chance accuracy.

If participants learned about the advisors in an unbiased manner, they would learn that the 75% advisor is just as likely to provide accurate advice as the 25% advisor is to provide inaccurate advice. However, if participants' learning was biased by optimistic initial beliefs and confirmation bias, their estimates of advisors' expertise would be inflated. In this case, participants would be more likely to follow the advice of the 75% advisor than to go against the advice of the 25% advisor. To compare the relative influence of the 75% advisor and the 25% advisor, we ran a second mixed model logistic regression with an advisor by advice interaction that directly tested whether the 75% advisor had a stronger influence on participants' choices than the 25% advisor.

Robust Bayesian estimation. For statistical comparisons, we adopted a Bayesian estimation approach that computes the posterior distribution of parameter estimates from the data. Specifically, the statistic of interest was assumed to be described by a t-distribution, with posterior estimates of the mean (μ), standard deviation (σ), normality (ν) estimated from the data using Markov chain Monte Carlo (MCMC) with noncommittal priors (see supplemental materials). We defined the 95% highest density interval (HDI) of the posterior distribution of μ as the 95% credible interval of μ. We refer to a "credible effect" whenever the 95% HDI does not include the comparison value. For example, if the 95% HDI of the posterior distribution of the mean within-participant difference in corrected average likelihood per trial (e.g., $\mu_{CB\text{-}BL}$) of two models does not include 0, we can conclude that that the corrected average likelihood per trial of the two models are credibly different. Effect size was defined as $\frac{\mu - 0}{\sigma}$. The Bayesian estimation approach and its advantages of over traditional t tests are fully described in Kruschke (2013).

Additional comparison models. We also compared the performance of the Confirmation Bias and Bayesian Learning models against other plausible models: a "Win-Stay-Lose-Shift" (WSLS) model that makes the same bet after a correct bet and the opposite bet after an incorrect bet, a reinforcement learning model with a temporal difference learning rule (TD), and a baseline "Null" model that fits a constant $p(bet = FOR)$ to each participant (see *Additional comparison models* in supplemental materials).

Data and code availability. Data for all three experiments, example stimuli, and custom code for all computational models are available at https://github.com/ycleong/AdviceTaking.

Results

Stock prediction phase. Participants correctly predicted the stock's price fluctuation on 55% of the trials, which was credibly higher than chance (95% HDI [52%, 58%]). As stock performance is stochastic (i.e., the price of the stock can decrease even when the trend suggests that the price of the stock is likely to increase), this measure likely underestimates participants' learning of the stock trend. We can use the Bayesian Learning model to obtain the best estimate of the stock trend given past outcomes of the stock. Even if participants' predictions were always consistent with the best estimate of the stock trend, they would only have been accurate on 60% of the trials (95% HDI [57%, 63%]).

For this phase of the task, we were interested in how participants predicted the stock rather than how *well* participants predicted the stock. In particular, we were interested in whether participants used the stock trend to make their stock predictions. We found that, on average, participants' predictions aligned with the stock trend on 66% of the trials (95% HDI [62%, 71%]). A mixed effects logistic regression model indicated that the stock trend reliably predicted participants' stock predictions (β = 0.95, 95% CI [0.86, 1.05], z = 18.8, $p < .001$), suggesting that participants relied on the stock's past performance when making their stock predictions.

Advisor evaluation phase. On each trial, participants bet for or against an advisor's prediction. These bets were indicative of participants' beliefs about how accurate they thought each advisor was on each trial. Over time, participants became more likely to bet for the 75% advisor's prediction (*M* = 84%, 95% HDI [79%, 93%]), indicating that they learned that this advisor was accurate. Participants also became less likely to bet for the 25% advisor's prediction (*M* = 23%, 95% HDI [12%, 29%]), indicating that they learned that this advisor was inaccurate. On average, participants bet for the 50% advisor on more than 50% of the trials (*M* = 67%, 95% HDI [60%, 78%]), indicating that they were optimistically biased in their assessment of this advisor (Figure 3A and 3B).

The 75% advisor was as accurate as the 25% advisor was inaccurate. If participants learned about the advisors in an unbiased manner, they would learn that the 75% advisor provides the same amount of useful information as the 25% advisor, and that they should bet for the former as much as they bet against latter. Instead, we found that participants were *more* likely to bet for the 75% advisor than they were to bet against the 25% advisor (M_{Diff} = 6%, 95% HDI [0%, 12%]), suggesting that their learning was optimistically biased. In the section below, we examine the optimism bias with model-based analyses.

Model-based analyses. The two computational models allow us to estimate participants' initial beliefs about the advisor's expertise from their bets on the advisor. We can then examine whether these beliefs are optimistically biased. Furthermore, the Bayesian Learning model and the Confirmation Bias model make different assumptions about how participants update their beliefs. By evaluating which model better fit participants' responses, we can assess which update rule best approximates participants' learning.

We fit both models to each participant's data to obtain the maximum a posteriori (MAP), or "best-fit," estimates of the model parameters (α, β, and τ) for that participant. The average best-fit

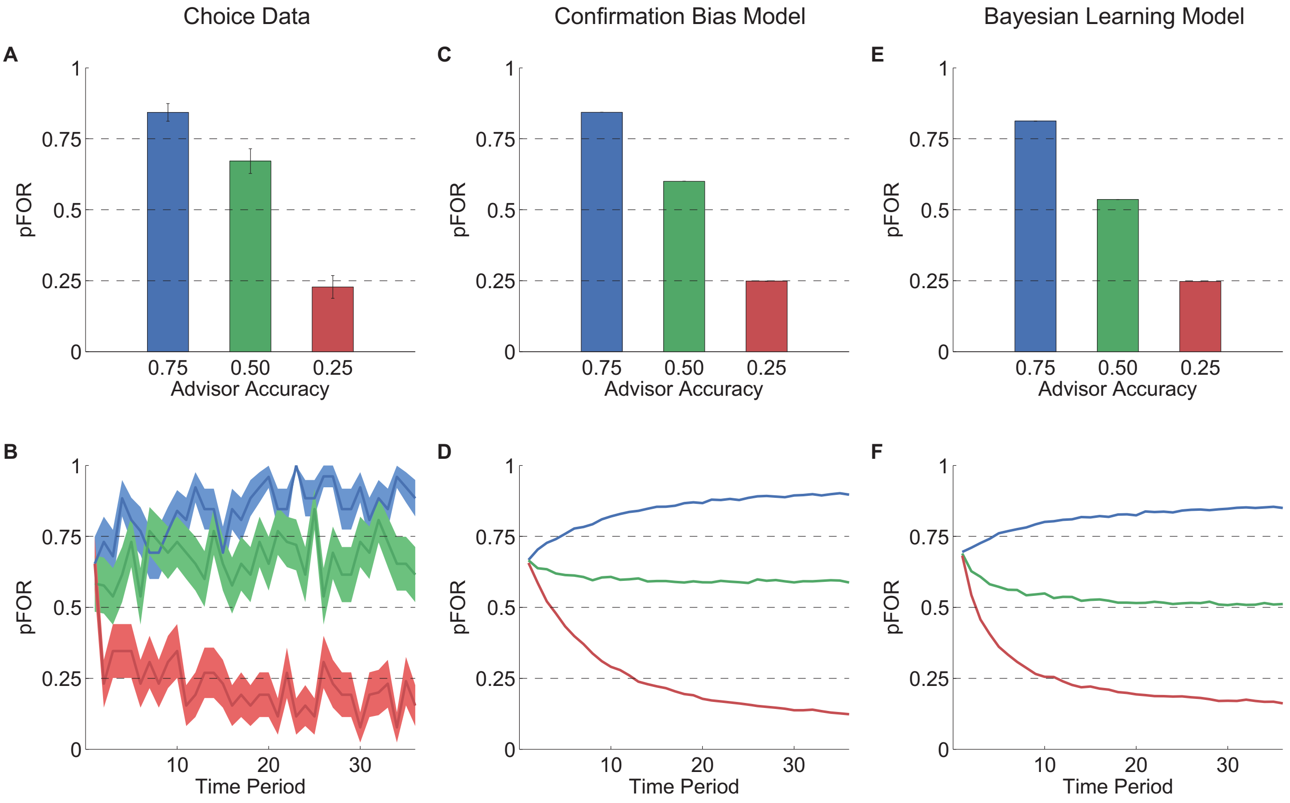

Figure 3. Comparison between participants' behavioral data and model simulations. On each time period, participants had to bet for or against an advisor's prediction. Participants performed repeated trials with 3 advisors, with accuracies of 75%, 50%, and 25% respectively. (A) Proportion of time periods on which participants bet for each advisor's prediction (pFOR), averaged across trials and (B) as a function of time period. Error bars and shading denote *SEM*. (C–D) Data generated by simulating the Confirmation Bias model with the best-fit model parameters over 500 iterations. (E–F) Data generated by simulating the Bayesian Learning model with the best-fit model parameters over 500 iterations. See the online article for the color version of this figure.

values of τ were 7.97 (*SE* = 0.84) and 11.23 (*SE* = 1.13) when participants' data were fit to the Confirmation Bias and Bayesian Learning model respectively (Fig. S5). A τ value in this range suggests that the average participant adopted a "soft-maximizing" choice strategy, such that the probability of betting for an advisor's prediction was weighted by the estimate of the advisor's accuracy (Fig. S4, see *Analysis of participants' betting strategies* in supplemental materials). This betting strategy is consistent with findings indicating that people rarely adopt a maximizing strategy (i.e., *always* choosing the option with higher probability of reward), despite it being the strategy that will earn them the most reward (Erev & Barron, 2005).

Next, we tested whether participants had optimistic initial beliefs about advisors' expertise. When we fit the models to participants' bets, both models fit participants' bets better when initialized with prior distributions that were optimistic (e.g., Figure 4C and 4D). That is, the best-fit values of α were often higher than that of β (Figure 4A and 4B). Given each participant's best-fit values of α and β, we can compute the mean estimate of that participant's initial beliefs about advisor expertise, $\hat{\alpha}_1$. For both models, $\hat{\alpha}_1$ was credibly above 0.5 (Confirmation Bias model: *M* = 0.62, 95% HDI [0.55, 0.70]); Bayesian Learning model: *M* = 0.59, 95% HDI [0.56, 0.65], Figure 4E and 4F), suggesting that optimistic initial beliefs indeed biased participants' learning. As the best-fit estimates of the initial beliefs could potentially be biased by later trials, we repeated this analysis fitting the models to only the first 12 trials of each advisor. The analysis yielded estimates that were near identical to that when the models were fit to all trials (Table S2).

We then asked whether participants exhibited confirmation bias when updating their beliefs by comparing the Bayesian Learning Model and the Confirmation Bias Model on how well they account for participants' bets. For each model, we computed the corrected average likelihood per trial, which denotes an unbiased estimate of the average likelihood of predicting a new data point given the model. We found that the Confirmation Bias model provided the better fit to most participants' data (19 of 26). Furthermore, the corrected average likelihood per trial was higher for the Confir-

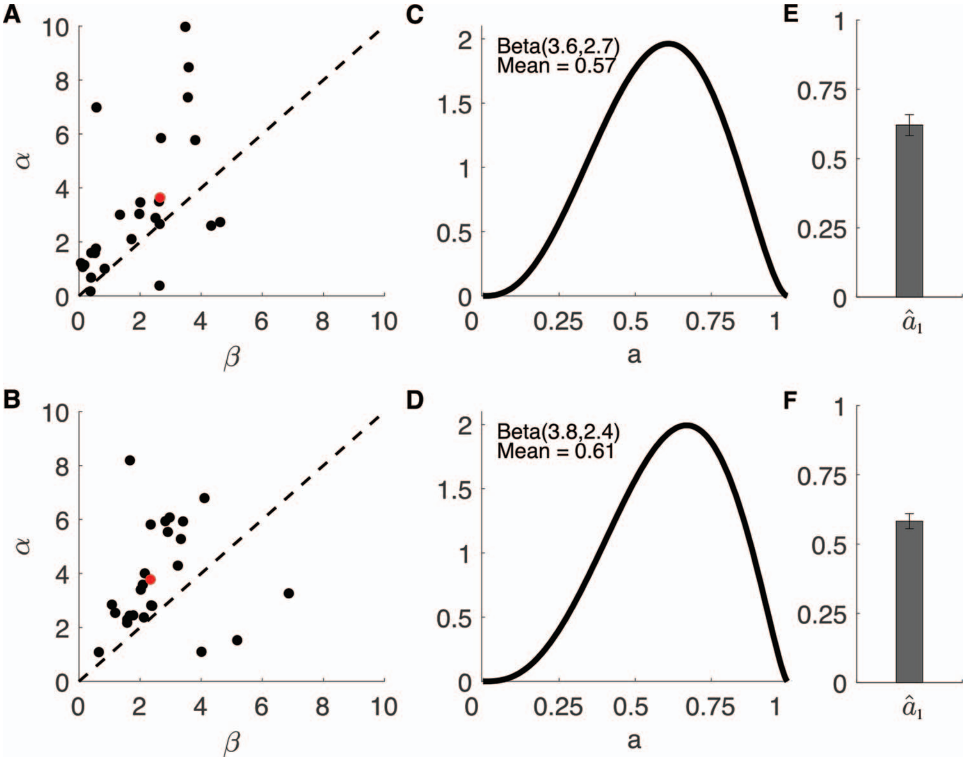

Figure 4. Majority of participants had optimistic priors about advisors' expertise. We fit the two models to find the MAP estimates of the model parameters for each participant. The MAP estimates of α and β are plotted on the left for (A) the Confirmation Bias Model, and (B) the Bayesian Learning Model. Each dot represents a participant. The black dashed line indicates values of α and β that give prior distributions that are not skewed (i.e., $\hat{\alpha}_1 = 0.5$). Points above the dashed line correspond to "optimistic priors" ($\hat{\alpha}_1 > 0.5$), whereas points below the dashed line correspond to "pessimistic priors" ($\hat{\alpha}_1 < 0.5$). For both models, the best-fit priors were optimistic for the majority of the participants. Plotted in the middle are the Beta distributions defined by (C) α = 3.6 and β = 2.7 and (D) α = 3.8 and β = 2.4. These illustrate the estimated prior distribution of a particular participant (red/grey dot in left panel) when fitting to the Confirmation Bias model and Bayesian Learning model respectively. Plotted on the right are the average estimates of $\hat{\alpha}_1$ for the (E) the Confirmation Bias Model and (F) the Bayesian Learning model. Error bars indicate *SEM*. See the online article for the color version of this figure.

mation Bias model than the Bayesian Leaning model ($M_{\text{CB-BL}}$ = 0.03, 95% HDI [0.01, 0.05]; Confirmation Bias Model: M = 0.65, 95% HDI [0.60, 0.70]; Bayesian Learning Model: M = 0.62, 95% HDI [0.58, 0.66]), with a medium effect size of 0.61 (95% HDI [0.16, 1.06]). The results from the model-based analyses are summarized in Table 1. These results indicate that the Confirmation Bias model provides a better account of participants' learning process, and suggest that participants weighted information consistent with their beliefs more than information inconsistent with their beliefs.

We then simulated the two models performing the *Advisor Evaluation phase* (Methods). These simulations provide us with the pattern of behavior we would have observed if participants' behavior was perfectly described by the models. We found that the simulated behavior of the Confirmation Bias model provided a better match to participants' data (Figure 3C and 3D). In particular, the Confirmation Bias model—like participants themselves—bet for the 50% advisor's predictions on more than 50% of the time periods, and bet for the 75% advisor's predictions more than they bet against the 25% advisor's predictions. In contrast, the Bayesian Learning model underestimated the bets for the 75% and 50% advisors. Furthermore, while the Confirmation Bias model exhibited persistent optimism in betting for the 50% advisor, the Bayesian Learning model is equally likely to bet for or against the 50% advisor after 20 time periods.

Taken together, our results suggest that both optimistic initial beliefs and confirmation bias contribute to learning optimistic estimates of advisors' expertise. Next, we examined whether these optimistic estimates biased subsequent advice taking in the *Joint Prediction Phase.*

Joint prediction phase. Results from the *Advisor Evaluation Phase* suggest that participants were biased in their estimates of advisors' accuracy at predicting the stock. Do these biases influence whether participants take advisors' recommendations when predicting the fluctuations of the stocks themselves? For each advisor, we calculated the proportion of trials on which participants followed or went against the advisor's recommendation. Participants did indeed follow the recommendation of the 50% advisor more often than chance (M = 68%, 95% HDI [0.59, 0.79]), providing further evidence that they overestimated the expertise of the 50% advisor and thought that he provided useful information. Participants were also more likely to follow the 75% advisor's recommendation than they were to go against the 25% advisor's recommendation (M = 7%, 95% HDI [3%, 10%]), demonstrating an optimistic bias in advice utilization that parallels the optimistic bias in learning about advisors.

Table 1
Experiment 1 Model-Fitting Results

	MAP estimate	
Parameter	Confirmation bias model	Bayesian learning model
α	3.23 (.51)	3.76 (.37)
β	1.96 (.28)	2.63 (.26)
τ	7.97 (.84)	11.23 (1.13)
$\hat{\alpha}_1$	.62 [.55, .70]	.61 [.55, .65]
	Model comparison	
AIC	94.5 (7.5)	103.6 (6.0)
# of best-fit participants	19	7
Corrected avg lik per trial	.65 [.60, .70]	.62 [.58, .66]
$\mu_{CB\text{-}BL}$	.03 [.01, .05]	
effect size	.61 [.16, 1.06]	

Note. We fit the two models to find the maximum a posteriori estimates (MAP) of the model parameters. For each participant, we fit the parameters that define their initial beliefs about the advisor (α and β), and the logistic function gain parameter τ. The table shows the average estimates of the model parameters. The mean estimate of $\hat{\alpha}_1$ was optimistic for both the confirmation bias model and the Bayesian learning model. We compared the two models based on AIC and the corrected average likelihood per Trial. For statistical inference, we used robust Bayesian estimation to estimate the mean difference in corrected average likelihood per trial of the two models ($\mu_{\text{CB-BL}}$). Also reported is the effect size when comparing $\mu_{\text{CB-BL}}$ to 0. Parentheses indicate standard error of the mean, while square brackets denote 95% HDI of the corresponding posterior distribution.

Participants could draw from two distinct sources of information when making their predictions—the trend of the stock based on its past performance (p), and the recommendation from an advisor. To maximize earnings in the task, participants should weight information based on relative accuracy. In particular, if an advisor was at chance at predicting the stock, the reward maximizing strategy would be to ignore that advisor's recommendation and rely solely on p. To examine how participants weight each piece of information, we ran a mixed effects logistic regression to predict participant's choices with p and the recommendations of each advisor as predictor variables. Trial-by-trial estimates of p were again generated by fitting the Bayesian Learning model to the stock outcomes. We found that p was a significant predictor of participants' predictions (β = 0.37, 95% CI [0.29, 0.45], z = 9.9, $p < .001$), suggesting that participants used the stock's past performance when making predictions. Participants positively weighted the advice of the 50% advisor (β = 0.45, 95% CI [0.39, 0.51], z = 15.2, $p < .001$). Given that the 50% advisor provided no useful information, participants should not have relied on this advisor. Their optimism in doing so was thus not only irrational, but materially costly in that it reduced their earnings in the task.

Participants positively weighted the recommendation of the 75% advisor (β = 1.42, 95% CI [1.32, 1.52], z = 28.7, $p < .001$) and negatively weighted the recommendation of the 25% advisor (β = −0.97, 95% CI [−1.05, −0.90], z = −26.0, $p < .001$), demonstrating that participants' weighting of advice was dependent on the expertise of the advisor. A second mixed model logistic regression found a significant Advisor x Advice interaction (β = 0.84, 95% CI [0.63, 1.01], z = 8.04, $p < .001$), indicating that participants were more likely to follow the 75% advisor's advice than they were to go against the 25% advisor's advice. Again, this reflects an irrational optimism bias in their perceptions of advisor expertise.

Explicit ratings. At the end of the experiment, participants estimated the percentage of time periods on which each advisor was accurate at predicting the stock. Participants' explicit ratings of the 75% and 50% advisor were indeed optimistic (75% advisor: M = 81%, 95% HDI [78%, 87%]; 50% advisor: M = 58%, 95% HDI [53%, 61%]) while their rating of the 25% advisor was not credibly different from 25% (M = 21%, 95% HDI [14%, 26%]).

Discussion

In Experiment 1, participants learned about the expertise of different financial advisors, and made predictions about the price fluctuation of a stock after hearing recommendations from the same advisors. While participants were sensitive to the *relative* accuracy of advisors and weighted advice accordingly, they were overly optimistic in their estimates of advisor's expertise, and relied on advisors more than warranted by the advisors' past performance. In particular, when presented with an advisor whose past accuracy was no better than chance, participants nonetheless believed that the advisor was meaningfully accurate and relied on this advisor's recommendations when making stock predictions themselves, even when doing so incurred a financial cost. Furthermore, participants credited an accurate advisor more than they penalized an equally inaccurate advisor, both in their beliefs about those advisors' accuracy, and in their willingness to take those advisors' recommendations when making predictions.

Using computational models, we demonstrated that participants' optimism reflected both optimistic initial beliefs and confirmation bias when updating those beliefs. Both components are necessary for the optimistic estimates to persist. If participants' initial beliefs were not optimistic, confirmation bias alone would not lead to systematically biased estimates. Similarly, if participants had optimistic initial beliefs but updated them in a statistically optimal manner, their estimates would converge on the advisor's true accuracy over time. Confirmation bias, however, meant that their optimistic initial beliefs were resistant to change despite repeated interactions with the advisor. These results are consistent with the other work demonstrating that expectancies influence impression formation (Bodenhausen, 1988; Hamilton, Sherman, & Ruvolo, 1990), visual perception (Summerfield & Egner, 2009), and even the sensation of pain (Atlas & Wager, 2012). Our study adds to this literature by breaking down expectancy effects into two separate components—initial expectations and confirmation bias, and formalizing each in a probabilistic model.

Our results might explain why "financial gurus" continue to attract a wide following and drive investment behavior, despite there being limited evidence that they can successfully predict the market (Engelberg, Sasseville, & Williams, 2012). Our results suggest that investors have inflated perceptions about the expertise of financial gurus, perceptions that are resistant to change due to confirmation bias. If a financial guru was highly inaccurate (a la our "25% advisor"), investors would likely catch on. However, if a financial guru has an accuracy that hovers around chance, optimistic initial beliefs and confirmation bias can lead investors to maintain excessive optimism about the guru's expertise.

Experiment 2

If optimistic initial beliefs indeed drive unrealistic optimism in advice-taking, manipulating participants' beliefs could mitigate or reverse these biases. In particular, our model suggests that participants who hold well-calibrated beliefs about advisors should not exhibit subsequent biases in their learning. In Experiment 2, we tested this prediction by manipulating participants' initial beliefs about two advisors who in fact performed at chance in predicting a stock's performance. We then observed how these expectations influenced participants' subsequent estimates of the advisors' expertise.

Method

Participants. Thirty participants were recruited from the Stanford community (14 male, 16 female, ages 19–49, mean age = 26.3). All participants provided written, informed consent prior to the start of the study. All experimental procedures were approved by the Stanford Institutional Review Board. Participants were paid up to $16 depending on their performance on the task.

Experimental task and design. Participants performed a variant of the Financial Advice Choice Task. Unlike in Experiment 1, participants encountered four (not three) financial advisors. They first completed 112 time periods of the *Stock Prediction* phase. Following this, they were introduced to the four financial advisors. Each advisor was associated with a star rating. Participants were told that these ratings were based on the advisors' past performance in making correct predictions about the stock. To ensure that participants remembered the ratings, we had them perform a recall task in which they were presented with the photograph of an advisor and had to indicate the associated rating. Participants had to make a correct response on each advisor for 8 consecutive trials before they could proceed to the *Advisor Evaluation* phase. Participants performed 112 time periods (28 with each advisor) of the *Advisor Evaluation* phase. The 4-star advisor was accurate on 75% of the time periods while the 1-star advisor was accurate on 25% of the time periods. The 3-star and 2-star advisors were both accurate on 50% of the time periods. Participants then performed 112 time periods of the *Joint Prediction* phase, in which they made predictions about the stock with a recommendation from one of the advisors (28 time periods with each advisor). At the end of the experiment, participants' were asked how accurate (0 −100%) they thought each advisor was at making predictions about the stock. Face stimuli were drawn from the same database as Experiment 1 and consisted of male Caucasian faces posing calm expressions with mouth closed and eyes gazing straight ahead.

Learning about advisors' expertise. Similar to Experiment 1, we calculated the proportion of time periods on which participants bet *for* each advisor's predictions in the *Advisor Evaluation* phase. To ensure that participants were able to distinguish between the most accurate advisor and the least accurate advisor, we first examined whether they bet for the 4-star advisor more than the 1-star advisor. Following this, we examined whether they bet for the 3-star advisor more than the 2-star advisor. The 3-star advisor and 2-star advisors were both at chance at predicting the stock, such that any differences in participants' evaluations of them were due to biased initial beliefs about each advisor, not their actual performance.

Computational modeling. We modeled participants' bets in the *Advisor Evaluation* phase using both the Bayesian Learning model and the Confirmation Bias model. The model fitting procedure was identical to Experiment 1, except that instead of fitting one initial belief distribution for all advisors, we fit a separate belief distribution for each of the four advisors. That is, we fit separate values of α and β for each advisor, reflecting that participants have different initial beliefs about each advisor. τ remains a participant-specific free parameter (i.e., one value for each participant). Again, we compared the model fits of the Bayesian Learning model and the Confirmation Bias model based on each model's corrected average likelihood per trial. Using the best-fitting parameters, we simulated the models performing the *Advisor Evaluation* phase. For more details about the models and model-fitting procedure, see the Methods section for Experiment 1.

Analysis of the Joint Prediction phase. In the *Joint Prediction* phase, we examined how participants weighted each advisor's recommendation when making stock predictions. We first ran a mixed effects logistic regression model that predicted participants' stock predictions from p and the recommendation from each advisor. We then ran a second regression to specifically test whether participants weighted the 3-star advisor's recommendation more than the 2-star advisor's recommendation. In both the Stock Prediction Phase and the Joint Prediction Phase, we applied the Bayesian Learning model to estimate p from the history of stock outcomes.

Results

Stock prediction phase. Participants correctly predicted the price fluctuation of the stock on 56% (95% HDI [53%, 58%]) of the time periods. On average, their predictions were consistent with the stock trend on 67% of the time periods (95% HDI [63% 69%]). The stock trend predicted participants' predictions (β = 0.97, 95% CI [0.86, 1.05], z = 19.1, $p < .001$), suggesting again that participants were able to track the stock trend when making their predictions.

Advisor evaluation phase. Participants tended to bet for the predictions of the 4-star advisor, who was accurate on 75% of the time periods (M = 90%, 95% HDI [85%, 100%]), and against the predictions of the 1-star advisor, who was accurate on 25% of the time periods (M = 20%, 95% HDI [10%, 27%]). The 3-star and 2-star advisors were equally accurate on the task (50% accuracy), but participants bet for the 3-star advisor's predictions credibly more than they did for the 2-star advisor's predictions (M_{Diff} = 34%, 95% HDI [18%, 51%]). While participants bet on 3-star advisor's prediction on more than 50% of the time periods (M = 76%, 95% HDI [69%, 84%]), their bets on the 2-star advisor's prediction were not credibly different from chance (M = 42%, 95% HDI [31%, 52%]). These results suggest that participants' initial beliefs biased their estimates of the advisors' expertise (Figure 5A and 5D).

To examine how these beliefs interacted with participants' learning, we fit the Bayesian Learning and Confirmation Bias models to participants' bets. Unlike in Experiment 1, we fit a different initial belief distribution to each advisor (i.e., different values for α and β), reflecting our prediction that the star ratings would bias participants' initial beliefs. As was the case in Experiment 1, the Confirmation Bias model was the better-fitting model for the majority of the participants (27 out of 30). The corrected average likelihood per trial of the Confirmation Bias model was

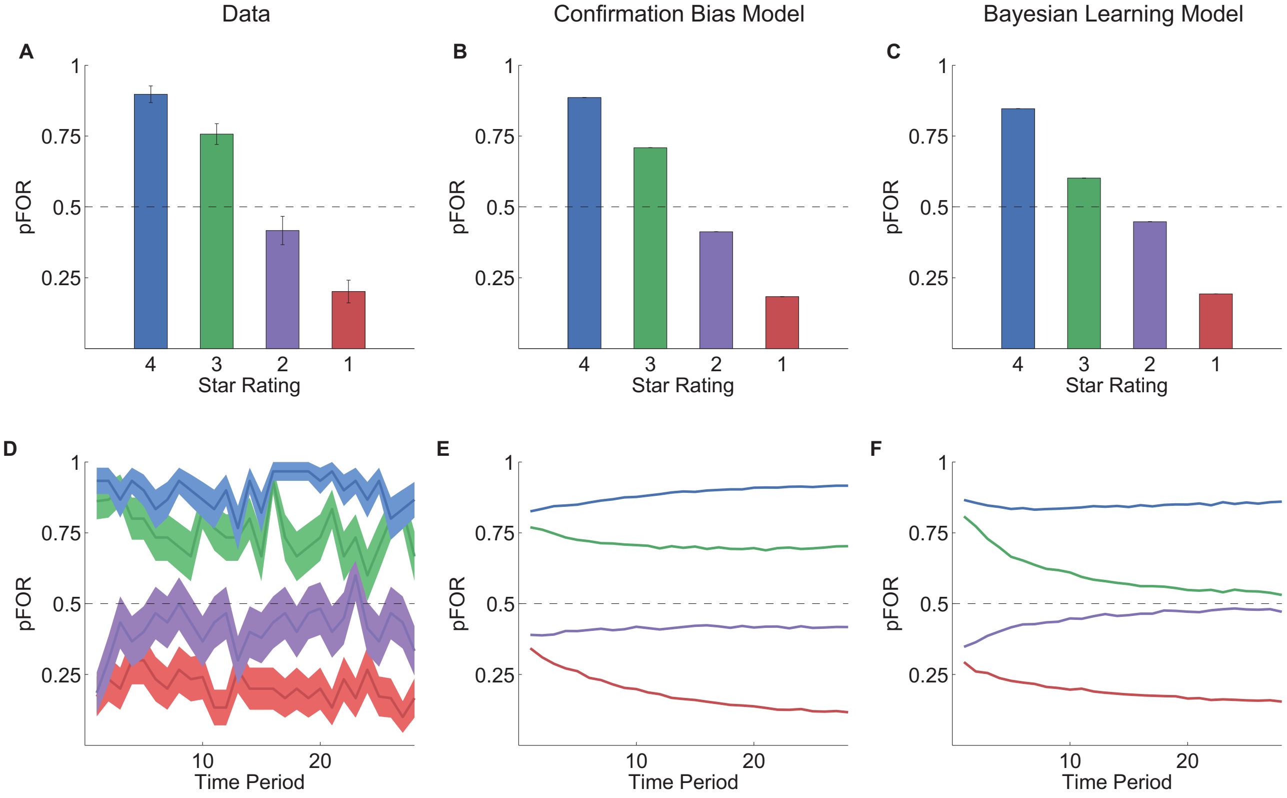

Figure 5. Experiment 2 Results. (A) Proportion of time periods on which participants bet for each advisor's prediction (pFOR), averaged across trials and (B) as a function of time period. Error bars and shading denote SEM. (C–D) Data generated by simulating the Confirmation Bias model with the best-fit model parameters over 500 iterations. (E–F) Data generated by simulating the Bayesian Learning model with the best-fit model parameters over 500 iterations. See the online article for the color version of this figure.

higher than that of the Bayesian Learning model (M_{Diff} = 0.04, 95% HDI [0.03, 0.06]; Confirmation Bias Model: M = 0.64, 95% HDI [0.59, 0.69]; Bayesian Learning Model: M = 0.59, 95% HDI [0.56, 0.63]), with a large effect size of 0.93 (95% HDI [0.49, 1.39]). When we simulated the models performing the task (Figure 5B, 5C, 5E, and 5F), the behavior of the Confirmation Bias Model matched participants' data better than that of the Bayesian Learning model. Despite being simulated with the best-fit initial belief distribution, the Bayesian Learning model underestimated participants' bets for the 3-star advisor and overestimated participants' bets for the 2-star advisor.

The best-fitting model parameters of the Confirmation Bias model suggest that participants were optimistic about the 4-star (M = 0.72, 95% HDI [0.70, 0.80]) and 3-star advisors (M = 0.69, 95% HDI [0.64, 0.76]), and pessimistic about the 1-star advisor (M = 0.43, 95% HDI [0.36, 0.48]). Participants' initial expectations about the 2-star advisor were not credibly different from chance (M = 0.45, 95% HDI [0.36, 0.53]), and were credibly lower than that for the 3-star advisor (M_{Diff} = −0.25, 95% HDI [−0.37, −0.14]). The results from the model-based analyses are summarized in Table 2, and were again very similar when the models were fitted to just the first 12 trials with each advisor (Fig. S2). Together, our results suggest that the star ratings biased participants' initial expectations about each advisor's expertise, which then led to differences in how they subsequently learned about the advisors.

Table 2
Experiment 2 Model-Fitting Results

Parameter	MAP estimate	
	Confirmation bias model	Bayesian learning model
4-star α	4.22 (.15)	4.90 (.22)
4-star β	1.69 (.18)	1.86 (.15)
4-star $\hat{\alpha}_1$	.72 [.70, .80]	.73 [.71, .79]
3-star α	4.21 (.19)	5.57 (.39)
3-star β	1.83 (.17)	2.60 (.12)
3-star $\hat{\alpha}_1$	.69 [.64, .76]	.67 [.63, .71]
2-star α	2.64 (.29)	3.26 (.21)
2-star β	3.32 (.31)	4.55 (.42)
2-star $\hat{\alpha}_1$	.45 [.36, .53]	.44 [.36, .50]
1-star α	2.95 (.33	2.89 (.27)
1-star β	3.64 (.23)	4.18 (.26)
1-star $\hat{\alpha}_1$	.43 [.36, .48]	.41[.33, .47]
τ	8.4 (.74)	11.11 (1.06)
	Model comparison	
AIC	103.4 (7.9)	118.9 (6.2)
# of best-fit participants	27	3
Corrected avg lik per trial	.64 [.59, .69]	.59 [.56, .63]
$\mu_{CB\text{-}BL}$	.044 [.025, .063]	
Effect Size	.929 [.487, 1.386]	

Note. We fit the two models to find the MAP estimates of the model parameters. For each participant, we fit the parameters that define their initial beliefs about each advisor (α and β), and the logistic function gain parameter τ. The table shows the average estimates of the model parameters. We compared the two models based on AIC and the corrected average likelihood per trial. For statistical inference, we used robust Bayesian estimation to estimate the mean difference in corrected average likelihood per trial of the two models ($\mu_{\mathrm{CB\text{-}BL}}$). Also reported is the effect size when comparing $\mu_{\mathrm{CB\text{-}BL}}$ to 0. Parentheses indicate standard error of the mean, while square brackets denote 95% HDI of the corresponding posterior distribution.

Joint prediction phase. Participants followed the recommendations of the 3-star advisor (M = 74%, 95% HDI [67%, 80%]) more often they followed the recommendations of the 2-star advisor (M = 48%, 95% HDI [38%, 58%]; M_{Diff}: 26%, 95% HDI [11%, 39%]). We ran a mixed effects logistic regression to examine how participants weighted their own estimate of the stock trend and the recommendations from each advisor when making their predictions. The stock trend was positively associated with their predictions (β = 0.52, 95% CI [0.44, 0.62], z = 12.15, $p < .001$), as was the recommendation of the 4-star advisor (β = 1.04, 95% CI [0.93, 1.17], z = 18.76, $p < .001$). In contrast, the 1-star advisor's recommendation was negatively associated with participants' predictions (β = −0.50, 95% CI [−0.58, −0.42], z = −12.5, $p < .001$). The 3-star advisor's recommendation was also positively associated with participants' predictions (β = 0.55, 95% CI [0.48, 0.63], z = 13.52, $p < .001$), but the 2-star advisor's recommendation was not (β = −0.04, 95% CI [−0.11, 0.33], z = −1.12, p = .263). In other words, participants relied on the 3-star advisor's recommendations when making their stock predictions, while ignoring the 2-star advisor's recommendation, even though both advisors had the same (chance) accuracy.

Explicit ratings. At the end of the experiment, participants were asked to estimate the percentage of time periods on which each advisor was accurate at predicting the stock. Even though the 3-star and 2-star advisors were both at chance at predicting the stock throughout the Advisor Evaluation and Joint Prediction phases, participants rated the 3-star advisor as being better than chance (M = 57.4%, 95% HDI [53%, 62%]), but the 2-star advisor as being not different from chance (M = 50.4%, 95% HDI [45%, 55%]).

Discussion

In Experiment 2, we manipulated participants' initial beliefs about an advisor's expertise by providing them with star ratings that supposedly reflected the advisor's past success in predicting the stock. Using computational models, we showed that differences in initial beliefs about advisors' expertise exerted persistent downstream effects on how participants learned about advisors and how they weighed their recommendation when making stock predictions. In particular, participants perceived a 3-star advisor as more accurate, and were more willing to utilize his advice, than a 2-star advisor, when in fact both advisors performed at chance.

These results highlight the contribution of optimistic initial beliefs to the unrealistic optimism in advice taking. Specifically, we found evidence of overly optimistic advice taking when participants were led to have optimistic initial beliefs (i.e., the 3-star advisor), but not when participants' initial beliefs were less optimistic (i.e., the 2-star advisor). Replicating the results from Experiment 1, the Confirmation Bias model provided a better fit to participants' bets than the Bayesian Learning model. Despite starting with different initial beliefs for the 3-star and 2-star advisors, the Bayesian Learning model's estimate for both advisors were roughly the same by the end of the *Advisor Evaluation* phase. This means that, if participants had updated their beliefs about the advisor's expertise in a Bayesian optimal manner, their experience with the advisor would have "overwritten" their initial beliefs, allowing them to arrive at an accurate estimate of each advisor's expertise. In contrast, in the Confirmation Bias model—and in participants' actual performance—initial beliefs persisted throughout the task, resulting in different estimates of expertise between the 3-star and 2-star advisors.

One heartening implication of the current results is that unrealistic optimism in advice-taking need not always occur. If participants have well-calibrated beliefs about an advisor, they will arrive at a relatively accurate estimate of the advisors' expertise, and weight their advice accordingly. Unfortunately, in most real-world decision-making scenarios, advisors are incentivized to enhance their perceived expertise to maximize the likelihood that decision-makers will heed their advice. How then, can decision-makers obtain an accurate expectation of advisor's expertise? One potential solution is to rely on crowd-sourced ratings, such as those used by Yelp and Amazon, as a prior on an advisor's expertise. One might expect that past decision-makers' estimates would converge on an advisors' true expertise, allowing future decision-makers to avoid optimism bias when encountering that advisor. Alternatively, if past decision-makers retain overly rosy assessments of advisors, their feedback could "spread" optimism bias to future decision-makers. In Experiment 3, we examined these possibilities by examining expertise judgments across "generations" of decision-makers.

Experiment 3

Websites such as Yelp, TripAdvisor, Rotten Tomatoes, and Amazon aggregate user ratings on restaurants, hotels, movies, products, and services. In recent years, several websites have begun to adopt a similar system to rate financial service providers. For example, WalletHub.com allows users to give a star rating to financial advisors based on their experience with them. These websites hope to arrive at accurate estimates by pooling over many individual estimates. In Experiment 3, we examined how such crowd-sourced ratings influence participants' learning about an advisor's expertise and utilization of the advisor's advice. A first group of participants performed the three phases of the Financial Advice Choice task and provided ratings on how accurate each advisor was. These ratings were then averaged and presented to a second group of participants prior to performing the same task. Previous work suggests that when aggregating estimates over many individuals, the average estimate can be remarkably close to the true values (Galton, 1907; Zarnowitz, 1984), a phenomenon known as "the wisdom of the crowd" (Surowiecki, 2005). If aggregated crowd-source ratings generate an accurate estimate of advisors' expertise, providing these ratings to subsequent decision-makers could help them calibrate their initial expectations about advisors and avoid optimism bias. However, if the crowd-source ratings are optimistically biased, they could propagate that bias to subsequent decision-makers. This would be akin to the phenomenon of "information cascades," in which decision-makers' choices are biased by having observed the choices of other decision-makers (Anderson & Holt, 1997).

Method

Participants. We recruited 100 participants (58 male, 41 female, 1 did not indicate sex, ages 19 to 61, mean age = 32.56) on Amazon Mechanical Turk (AMT). Fourteen participants were excluded for missing more than 10% of the trials ($n = 4$), failing an attention check question ($n = 9$) or for only pressing one button throughout the task ($n = 1$). We call this group *Generation 1*. We recruited a second group of 100 participants on Amazon Mechanical Turk (52 male, 46 female, 2 did not indicate sex, ages 19–69, mean age = 33.26). Eight of these participants were excluded for missing more than 10% of the trials ($n = 4$) or for failing the attention check question ($n = 4$). We call this second group *Generation 2*. For both generations, the duration of the task was around 20 min. Participants were compensated $1.00 for their time and could earn a bonus of up to $3.50 depending on their performance on the task. We increased the planned sample size considerably due to the smaller number of trials, and because prior experience with learning experiments on the AMT platform suggested to us that there would be a larger proportion of the sample who would fail to learn the task. To ensure data quality, we recruited only participants with approval ratings (i.e., HIT approval ratings) of greater than 95% on the Amazon Mechanical Turk interface. All experimental procedures were approved by the Stanford Institutional Review Board.

Generation 1. Participants performed a shorter version of the FAC task, in which there were 36 time periods in the *Stock Prediction* phase, 60 time periods in the *Advisor Evaluation* phase (20 with each advisor), and 60 time periods in the *Joint Prediction* phase (20 with each advisor). As in Experiment 1, participants encountered three advisors—one who was 75% accurate, one who was 50% accurate and one who was 25% accurate. For each participant, the photo representing each advisor was randomly selected from a stimulus set of 18 photos. Face stimuli were drawn from the same database as Experiment 1 and consisted of male Caucasian faces posing calm expressions with mouth closed and eyes gazing straight ahead. At the end of the task, participants were presented with six photos, and were told to rate the three advisors that they encountered in the task. Participants could give each advisor a star rating ranging from 1 star to 5 stars (integer values only). Data from participants who did not rate the three advisors they encountered or who rated more than or less than 3 advisors were discarded. Data from Generation 1 provided us with an opportunity to examine a direct replication of Experiment 1. We analyzed the data using the same procedures described in Experiment 1.

Generation 2. Participants performed the same FAC task as generation 1. However, prior to the start of the *Advisor Evaluation* phase, participants were presented with a star rating for each of the three advisors. These were the average ratings of a 75% advisor, a 50% advisor, and 25% made by Generation 1 participants. As before, photographs were randomly paired with specific accuracy levels. Participants were instructed to remember the ratings as they would be quizzed on them shortly. Following this, participants were given a memory test where they were presented with the photo of one of the advisors and a choice of three possible ratings associated with the advisor. If participants chose the correct rating, they would move on to the next trial. If participants chose an incorrect rating, they would be told that they were incorrect, and given the correct answer before moving on to the next trial. Participants performed 9 trials of the memory test (3 trials with each advisor in a randomized order). Finally, participants performed an attention check question, in which they were presented with the three advisors and had to match each advisor to a rating. Participants who failed to correctly match the ratings to the corresponding advisor were allowed to proceed with the task, but their data were subsequently discarded. At the end of the experiment, participants were presented with photos of six advisors, and had to rate, from 1–5 stars, the three advisors whom they had encountered. Data from participants who did not rate the three advisors they encountered or who rated more than or less than 3 advisors were discarded.

The setup for generation 2 is similar to that in Experiment 2 reported above. In Experiment 2, we experimentally manipulated two advisors with the same (chance) accuracy to have different star ratings. In generation 2, the ratings were not experimenter-generated but were instead obtained by averaging the ratings of participants who had previously performed the task. Nevertheless, generation 2 provided us with the opportunity to replicate our findings in Experiment 2 that manipulating participants' initial expectations about an advisor can bias participants' learning about the advisor. Data from generation 2 were analyzed using the same procedures described in Experiment 2.

Results

Generation 1. Data from Generation 1 replicated all key findings from Experiment 1. In the *Stock Prediction* phase, participants' stock predictions were reliably predicted by the stock trend ($\beta = 0.47$, 95% CI [0.16, 0.38], $z = 12.1$, $p < .001$). In the *Advisor Evaluation* phase, participants bet for the 50% advisor's prediction on more than 50% of the time periods ($M = 58\%$, 95% HDI [54%, 64%]), suggesting that participants overestimated the advisor's expertise (Figure 6A and 6B). Participants also bet for the 75% advisor more than they bet against the 25% advisor, providing further evidence of an optimism bias when learning about advisors ($M_{\mathrm{Diff}} = 9\%$, 95% HDI [5%, 13%]). The Confirmation Bias model again provided a better fit to participants' data than the Bayesian Learning model, and the best-fitting initial belief distribution for both models were optimistic (Table S3).

In the *Joint Prediction* phase, participants followed the recommendation of the 50% advisor on more than 50% of the time periods ($M = 60\%$, 95% HDI [55%, 64%]), and followed the 75% advisor's recommendation more often than they went against the 25% advisor's recommendation ($M = 18\%$, 95% HDI [14%, 22%]). A mixed effects logistic regression provided further evidence that participants weighted the 50% advisor's recommendation when making their own predictions ($\beta = 0.40$, 95% CI [0.29, 0.50], $z = 8.06$, $p < .001$), and relied on the 75% advisor's

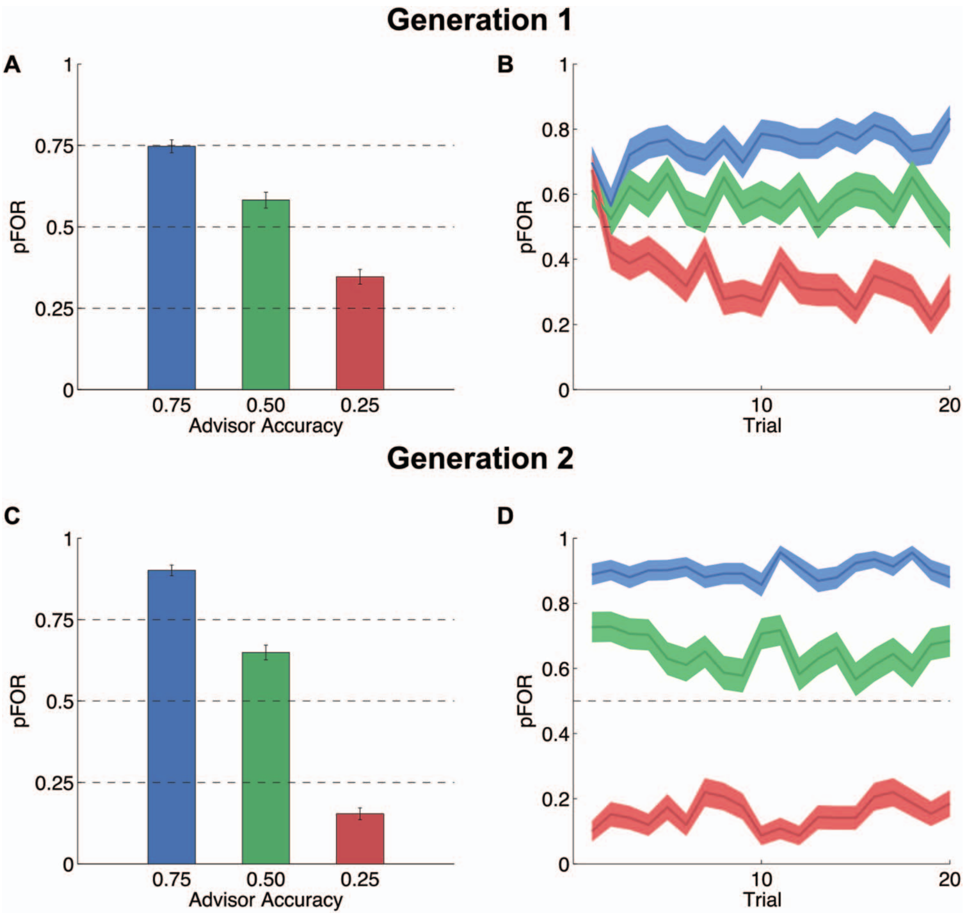

Figure 6. Experiment 3 Results. (A) Proportion of time periods on which Generation 1 participants bet for each advisor's prediction (pFOR), averaged across trials and (B) as a function of time period. Error bars and shading denote SEM. (C–D) Results from Generation 2 participants. See the online article for the color version of this figure.

recommendation more than they went against the 25% advisor's recommendation (β = 1.8, 95% CI [1.5, 2.2], z = 11.7, p < .001).

At the end of the experiment, we asked participants to give each of the advisors a star rating (out of 5 stars). On average, the 75% advisor was rated 3.95 stars (95% HDI [3.78, 4.13]), the 50% advisor was rated 2.97 stars (95% HDI [2.79, 3.14]), and the 25% advisor was rated 1.98 stars (95% HDI [1.75, 2.15]).

Generation 2. Participants in Generation 2 were provided with the average ratings of each advisor prior to the start of the *Advisor Evaluation* phase. On average, participants bet for the 50% advisor's prediction credibly more than 50% of the time periods (M = 65%, 95% HDI [60%, 70%]), suggesting that similar to participants from Generation 1, participants in Generation 2 were optimistic about this advisor's expertise. Participants bet for the 75% advisor's prediction on 90% (95% HDI [88%, 95%]) of the time periods, and bet for the 25% advisor's prediction on 15% (95% HDI [5%, 20%]) of the time periods (Figure 6C and 6D). We then fit the Confirmation Bias model to participants' bets to estimate participants' initial beliefs about each advisor to examine the influence of the star ratings (Table S4). The 2.97 star rating of the 50% advisor led to an optimistic initial expectation about the advisor's expertise (M = 0.63, 95% HDI [0.60, 0.66]), suggesting that the star ratings propagated optimism about the 50% advisor from Generation 1 to Generation 2. The 3.95 star rating of the 75% advisor and 1.98 star rating of the 25% advisor, respectively, led to optimistic (M = 0.78, 95% HDI [0.76, 0.79]) and pessimistic initial expectations (M = 0.35, 95% HDI [0.32, 0.38]). The initial expectations of the 75% advisor were also more extreme than those of the 25% advisor (M_{Diff} = 0.10, 95% HDI [0.07, 0.13]).

In the *Joint Prediction* phase, participants weighted the 50% advisor's recommendation positively when making their predictions (β = 0.70, 95% CI [0.60, 80], z = 13.8, p < .001), indicating that participants were again utilizing the advice of an advisor that provided no useful information. Participants weighted the 75% advisor's recommendation positively more than they weighted the 25% advisor's recommendation negatively (β = −2.1, 95% CI [−2.46, −1.72], z = −11.1, p < .001), again reflecting an optimistic bias in advice utilization.

At the end of the experiment, participants gave the 75% advisor a rating of 3.88 (95% HDI [3.74, 4.04]), the 50% advisor a rating of 2.68 (95% HDI [2.56, 2.84]) and the 25% advisor rating of 1.65 (95% HDI [1.45, 1.77]).

Discussion

Decision-makers increasingly rely on crowd-sourced ratings when making decisions. In Experiment 3, we explored the influence of crowd-sourced ratings on perceptions of advisors' expertise in our task. One group of participants performed the task and provided ratings for each of the advisors. These ratings were averaged and passed to a second group of participants who then performed the same task. We were particularly interested in investigating if the aggregated ratings would help correct for the optimistic initial beliefs we observed in Experiment 1, such that the second group of participants would start the task with well-calibrated expectations and thus not exhibit excessive optimism in their perceptions of the 50% advisor.

We had the first group of participants rate the advisors on a 1–5 star rating scale commonly used on websites that aggregate user ratings (e.g., Yelp, Amazon, Wallethub). Although we do not know how participants would map the star ratings to specific levels of advisor accuracy, we can measure the effect of the ratings by using the Confirmation Bias model to estimate the initial expectations of the second group of participants. We found that the second group of participants had optimistic initial expectations of the 75% advisor, but pessimistic initial expectations of the 25% advisor, indicating that the star ratings biased participants' initial expectations. Initial expectations of the 50% advisor were optimistic, indicating that instead of calibrating initial beliefs, crowd-sourced ratings propagated optimistic expectations to a second generation of participants. This finding is consistent with other work that have examined the factors that undermine the accuracy of aggregated estimates (Einhorn, Hogarth, & Klempner, 1977; Lorenz, Rauhut, Schweitzer, & Helbing, 2011). Aggregated estimates tend to be accurate only when there is little systematic bias in the individual estimates. In our task, the first group of participants, on average, exhibited optimistic initial beliefs about the advisors. Confirmation bias allowed that optimism to persist, despite repeated experience to the contrary, resulting in overly optimistic star ratings for the 50% advisor. These ratings then propagated and exaggerated optimism bias in decision-makers who received them.

A similar process is thought to underlie the formation of information cascades, a type of herding behavior that has been extensively studied in economics (Anderson & Holt, 1997; Bikhchandani, Hirshleifer, & Welch, 1992; Chamley, 2003). An information cascade occurs when individual decision-makers ignore their own private information in favor of following the actions of others. Whereas information cascades involve individual decision-makers making sequential decisions, our finding that the expectations engendered by the ratings from a previous group overwhelms one's own experience is reminisce of this process.

General Discussion

Decision-makers seek out advice because they believe that others have relevant information or expertise that could be useful to them (Harvey & Fischer, 1997). By combining what they know with what others tell them, decision-makers can make more accurate judgments (Bahrami et al., 2010; Yaniv, 2004), choose actions with greater rewards (Biele, Rieskamp, Krugel, & Heekeren, 2011; Li, Delgado, & Phelps, 2011), and avoid costly mistakes (Olsson & Phelps, 2007). Following advice can be beneficial, but only if the advice provides accurate and relevant information that can guide decisions. Instead of improving the quality of decisions, relying on bad advice can lead to inaccurate judgments (Bahrami et al., 2010), a diminished ability to learn from feedback, and perseveration of suboptimal behavior (Doll et al., 2009). To make adaptive decisions, one has to learn what advice to follow and what advice to ignore.

Although there has been a fair amount of research examining how decision-makers take into account advisors' expertise when making decisions (Bonaccio & Dalal, 2006; Budescu & Rantilla, 2000; Toelch, Bach, & Dolan, 2014; Toelch, Bruce, Newson, Richerson, & Reader, 2014), less is known about how decision-makers learn about advisors' expertise in the first place. Previous work has primarily modeled the tracking of advisor expertise as a Bayesian optimal process (Behrens et al., 2008; Boorman et al., 2013; Diaconescu et al., 2014; Shafto, Eaves, Navarro, & Perfors, 2012). Here, we document an optimism bias in how decision-makers learn about the expertise of a financial advisor and in how they decide to take those advisors' suggestions. By fitting and comparing computational models, we demonstrated that the optimism bias is attributable to the combination of optimistic initial beliefs and confirmation bias in how those beliefs are updated. Our work extends the existing literature in two ways. First, unlike previous work, we did not assume priors, but instead estimated them from the data. This afforded us the tools to identify biases in participants' initial beliefs. Second, we formalized a model that incorporated confirmation bias, such that expectation-consistent information is weighted more than expectation inconsistent information. Confirmation bias is a pervasive cognitive bias that has been shown in many domains (Nickerson, 1998; Oswald & Grosjean, 2004). Our work demonstrates that confirmation bias influences how decision-makers track the expertise of their advisors and has subsequent effects on advice-taking behavior.

Our model predicts that when decision-makers' initial expectations are well calibrated, their estimate of an advisor's expertise will remain fairly accurate. The results from Experiment 2 confirmed this prediction. We explicitly manipulated participants' expectations, and demonstrated that when participants had a less optimistic initial belief about an advisor, they correctly recognized an advisor as being at chance and did not utilize his advice in making their predictions. These results highlight the importance of calibrating decision-maker's initial beliefs. In Experiment 3, we investigated a popular method aimed at providing decision-makers with accurate expectations—aggregated crowd sourced ratings collected from past decision-makers. We found that although each decision-maker rated the advisors independently, their ratings were systematically biased. As a result, the aggregated ratings propagated and likely enhanced the optimism bias to future decision-makers.

Our task was set in a simplified and controlled setting—in which advisors made binary recommendations and decision-makers received immediate and explicit feedback about the advisor's performance. Would the current results generalize to more realistic decision-making scenarios, where advisors can embellish their recommendations and feedback is often delayed and complex? We argue that the effects of expectations and confirmation bias would be even stronger in such situations, as they leave greater cognitive flexibility for decision-makers to "explain away" the feedback to arrive at an expectation consistent conclusion (Hamilton et al., 1990). That is, decision-makers would be able to generate more excuses for why an advisor they had expected to be accurate made an inaccurate prediction. This hypothesis can be tested in future experiments, and would provide us with greater insights into the cognitive operations that lead to confirmation bias.

One likely reason why participants have optimistic initial expectations about advisor's expertise is that the advisors in our task were given the title of "financial advisor," and participants might have inferred from the title that the advisors had domain-specific expertise in predicting stocks. Expert advisors are often given credentials and described in years of experience, and a similar inference might explain optimism in expert advisors in domains beyond financial advice taking. Would participants be optimistically biased if the advisors had not been described as financial advisors? Our speculation is that they would, though perhaps to a smaller extent. A positivity bias has long been documented in the person perception literature, whereby target individuals are predominantly evaluated positively in the absence of clear positive or negative information (Bruner & Tagiuri, 1954; Klar & Giladi, 1997; Sears, 1983). This positivity bias has been found to influence the attribution of specific traits such as trustworthiness and competence, emerging as early as 100 ms of viewing a target face (Willis & Todorov, 2006), and is thought to have its origins early in childhood (Boseovski, 2010). For example, studies with 3- to 4-year-olds have found that although children are able to distinguish the relative accuracies of informants, there remains a robust bias to trust information provided by an informant who had been inaccurate in the past (Jaswal, Croft, Setia, & Cole, 2010; Vanderbilt, Heyman, & Liu, 2014). Having optimistic initial expectations of others is thought to be adaptive, as they encourage learning from others and facilitate cooperation (Baier, 1986; Boseovski, 2010; Hardin, 1993). In contrast, a decision-maker who has pessimistic initial expectations of advisors would rarely heed advice, and thus not benefit when an advisor has useful information to offer.

Unrealistic optimism has been previously documented in the context of an individual's prediction about future events (Shepperd, Waters, Weinstein, & Klein, 2015; Weinstein, 1980). In particular, people tend to overestimate the probability of positive events and underestimate the probability of negative events happening to them. These optimistic estimates persist even in light of new information indicating that they are miscalibrated (Armor & Taylor, 2002; Weinstein & Klein, 1995), much like the unrealistic optimism in advisors' expertise reported in this paper. Unrealistic optimism in oneself is thought to be similarly related to biased initial expectations and updating (Gerrard, Gibbons, & Reis-Bergan, 1999; Kuzmanovic & Rigoux, 2017; Sharot, Korn, & Dolan, 2011; Weinstein, 1987), suggesting that the phenomenon can be accounted for by the computational account presented in this paper.

In most real-world advice-taking scenarios, both optimism about oneself and optimism about others can lead to inflated perceptions of advisor expertise and overreliance on advice. For example, decision-makers may be particularly optimistic about the expertise of advisors who provide them with favorable information (e.g., telling the decision-maker that a stock that the decision-maker has a stake in will increase in value). In such cases, the optimistic expectations about the advisors could derive from optimism about gaining rewards as well as optimistic expectations about the expertise of advisors more broadly. In our task, participants were rewarded for correctly predicting the stock price fluctuation (Stock Prediction phase and Joint Prediction phase) or whether an advisor would be accurate in his prediction (Advisor Evaluation phase). As such, participants had no explicit motivation to *want* the advisors to be accurate. This was a deliberate design decision to allow us to disentangle optimism about others from optimism about the self. Optimism about others is a less studied phenomenon, and our study demonstrated that optimistic initial expectations about advisors' expertise can contribute to overreliance on advice, independent of optimism about the self.

The mechanisms underlying the optimistic bias observed in the current set of studies are basic cognitive processes that are likely to impact advice-taking behavior beyond financial decision-making scenarios. From health care professionals to political pundits, policy advisors to sports commentators, advisors are often portrayed as experts in their respective fields. Decision-makers are likely to have optimistic expectations about these advisors, expectations that could be wrong yet resistant to change. We believe that this research highlights the importance of tabulating and making public quantitative metrics of advisor accuracy, such that decision-makers can consider them when deciding whether to utilize a piece of advice. Advisors are often helpful, but knowing when they are not can help decision-makers discern how to incorporate advice when making choices.

Context

Across a variety of domains, people repeatedly rely on the advice of "experts" who are no better than chance at making accurate predictions. This behavior is puzzling, and incompatible with the popular view of human social learning as statistically optimal Bayesian inference. In this paper, we document an optimism bias in how people learn about expert advisors, and demonstrate that, with new modifications, existing models of learning can be extended to account for the observed behavior. In doing so, we provide an improved computational account of how people learn from and about people, and explored possible strategies to curb suboptimal advice taking. We hope to take the current work in three directions: (1) investigate the relationship between biased priors (as estimated by our computational models) and implicit attitudes (as measured by the Implicit Association Test and other implicit measures), (2) investigate the interaction between motivation and expectation in advice-taking (e.g., how does wanting an outcome influence the perception of an advisor offering favorable advice?) (3) use our computational models to generate trial-by-trial estimates of participants' learning and regress these estimates against neuroimaging data to search for neural correlates of biased updating.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19,* 716–723. http://dx.doi.org/10.1109/TAC.1974.1100705

Akaike, H. (1978). On the likelihood of a time series model. *The Statistician, 27,* 217–235. http://dx.doi.org/10.2307/2988185

Anderson, L. R., & Holt, C. A. (1997). Information cascades in the laboratory. *The American Economic Review, 87,* 847–862.

Armor, D. A., & Taylor, S. E. (2002). When predictions fail: The dilemma of unrealistic optimism. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 334–347). New York, NY: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511808098.021

Atlas, L. Y., & Wager, T. D. (2012). How expectations shape pain. *Neuroscience Letters, 520,* 140–148. http://dx.doi.org/10.1016/j.neulet.2012.03.039

Bahrami, B., Olsen, K., Latham, P. E., Roepstorff, A., Rees, G., & Frith, C. D. (2010). Optimally interacting minds. *Science, 329,* 1081–1085. http://dx.doi.org/10.1126/science.1185718

Baier, A. (1986). Trust and antitrust. *Ethics, 96,* 231–260. http://dx.doi.org/10.1086/292745

Baker, C. L., Saxe, R., & Tenenbaum, J. B. (2009). Action understanding as inverse planning. *Cognition, 113,* 329–349. http://dx.doi.org/10.1016/j.cognition.2009.07.005

Behrens, T. E. J., Hunt, L. T., Woolrich, M. W., & Rushworth, M. F. S. (2008). Associative learning of social value. *Nature, 456,* 245–249. http://dx.doi.org/10.1038/nature07538

Behrens, T. E. J., Woolrich, M. W., Walton, M. E., & Rushworth, M. F. S. (2007). Learning the value of information in an uncertain world. *Nature Neuroscience, 10,* 1214–1221. http://dx.doi.org/10.1038/nn1954

Berlo, D. K., Lemert, J. B., & Mertz, R. J. (1969). Dimensions for evaluating the acceptability of message sources. *Public Opinion Quarterly, 33,* 563–576. http://dx.doi.org/10.1086/267745

Biele, G., Rieskamp, J., & Gonzalez, R. (2009). Computational models for the combination of advice and individual learning. *Cognitive Science, 33,* 206–242. http://dx.doi.org/10.1111/j.1551-6709.2009.01010.x

Biele, G., Rieskamp, J., Krugel, L. K., & Heekeren, H. R. (2011). The neural basis of following advice. *PLoS Biology, 9,* e1001089. http://dx.doi.org/10.1371/journal.pbio.1001089

Bikhchandani, S., Hirshleifer, D., & Welch, I. (1992). A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of Political Economy, 100,* 992–1026. http://dx.doi.org/10.1086/261849

Birnbaum, M. H., & Stegner, S. E. (1979). Source credibility in social judgment: Bias, expertise, and the judge's point of view. *Journal of Personality and Social Psychology, 37,* 48–74. http://dx.doi.org/10.1037/0022-3514.37.1.48

Bodenhausen, G. V. (1988). Stereotypic biases in social decision making and memory: Testing process models of stereotype use. *Journal of Personality and Social Psychology, 55,* 726–737. http://dx.doi.org/10.1037/0022-3514.55.5.726

Bonaccio, S., & Dalal, R. S. (2006). Advice taking and decision-making: An integrative literature review, and implications for the organizational sciences. *Organizational Behavior and Human Decision Processes, 101,* 127–151. http://dx.doi.org/10.1016/j.obhdp.2006.07.001

Boorman, E. D., O'Doherty, J. P., Adolphs, R., & Rangel, A. (2013). The behavioral and neural mechanisms underlying the tracking of expertise. *Neuron, 80,* 1558–1571. http://dx.doi.org/10.1016/j.neuron.2013.10.024

Boseovski, J. J. (2010). Evidence for "rose-colored glasses": An examination of the positivity bias in young children's personality judgments. *Child Development Perspectives, 4,* 212–218. http://dx.doi.org/10.1111/j.1750-8606.2010.00149.x

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

Bruner, J. S., & Tagiuri, R. (1954). The perception of people. In G. Lindzey (Ed.), *Handbook of social psychology* (Vol. 2, pp. 634–654). Reading, MA: Addison Wesley.

Budescu, D. V., & Rantilla, A. K. (2000). Confidence in aggregation of expert opinions. *Acta Psychologica, 104,* 371–398. http://dx.doi.org/10.1016/S0001-6918(00)00037-8

Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: Understanding AIC and BIC in model selection. *Sociological Methods & Research, 33,* 261–304. http://dx.doi.org/10.1177/0049124104268644

Chamley, C. (2003). *Rational herds: Economic models of social learning*. New York, NY: Cambridge University Press. Retrieved from http://www.amazon.ca/exec/obidos/redirect?tag=citeulike09-20&path=ASIN/052153092X. http://dx.doi.org/10.1017/CBO9780511616372

Daw, N. D. (2011). Trial-by-trial data analysis using computational models. In E. A. Phelps, T. W. Robbins, & M. R. Delgado (Eds.), *Affect, learning and decision making, Attention and performance XXIII.* New York, NY: Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780199600434.003.0001

Diaconescu, A. O., Mathys, C., Weber, L. A. E., Daunizeau, J., Kasper, L., Lomakina, E. I., . . . Stephan, K. E. (2014). Inferring on the intentions of others by hierarchical Bayesian learning. *PLoS Computational Biology, 10,* e1003810. http://dx.doi.org/10.1371/journal.pcbi.1003810

Doll, B. B., Jacobs, W. J., Sanfey, A. G., & Frank, M. J. (2009). Instructional control of reinforcement learning: A behavioral and neurocomputational investigation. *Brain Research, 1299,* 74–94. http://dx.doi.org/10.1016/j.brainres.2009.07.007

Einhorn, H. J., Hogarth, R. M., & Klempner, E. (1977). Quality of group judgment. *Psychological Bulletin, 84,* 158–172. http://dx.doi.org/10.1037/0033-2909.84.1.158

Engelberg, J., Sasseville, C., & Williams, J. (2012). Market madness? The case of mad money. *Management Science, 58,* 351–364. http://dx.doi.org/10.1287/mnsc.1100.1290

Erev, I., & Barron, G. (2005). On adaptation, maximization, and reinforcement learning among cognitive strategies. *Psychological Review, 112,* 912–931. http://dx.doi.org/10.1037/0033-295X.112.4.912

Galton, F. (1907). Vox populi [voice of the people]. *Nature, 75,* 450–451. http://dx.doi.org/10.1038/075450a0

Gelman, A., Hwang, J., & Vehtari, A. (2014). Understanding predictive information criteria for Bayesian models. *Statistics and Computing, 24,* 997–1016. http://dx.doi.org/10.1007/s11222-013-9416-2

Gerrard, M., Gibbons, F. X., & Reis-Bergan, M. (1999). The effect of risk communication on risk perceptions: The significance of individual differences. *JNCI Monographs, 1999,* 94–100.

Hamilton, D. L., Sherman, S. J., & Ruvolo, C. M. (1990). Stereotype-based expectancies: Effects on information processing and social behavior. *Journal of Social Issues, 46,* 35–60. http://dx.doi.org/10.1111/j.1540-4560.1990.tb01922.x

Hardin, R. (1993). The street-level epistemology of trust. *Politics & Society, 21,* 505–529. http://dx.doi.org/10.1177/0032329293021004006

Harvey, N., & Fischer, I. (1997). Taking advice: Accepting help, improving judgment, and sharing responsibility. *Organizational Behavior and Human Decision Processes, 70,* 117–133. http://dx.doi.org/10.1006/obhd.1997.2697

Hovland, C. I., Janis, I. L., & Kelley, H. H. (1953). *Communication and persuasion; psychological studies of opinion change* (Vol. xii). New Haven, CT: Yale University Press.

Hurvich, C. M., & Tsai, C.-L. (1989). Regression and time series model selection in small samples. *Biometrika, 76,* 297–307. http://dx.doi.org/10.1093/biomet/76.2.297

Jaswal, V. K., Croft, A. C., Setia, A. R., & Cole, C. A. (2010). Young children have a specific, highly robust bias to trust testimony. *Psychological Science, 21,* 1541–1547. http://dx.doi.org/10.1177/0956797610383438

Klar, Y., & Giladi, E. E. (1997). No one in my group can be below the group's average: A robust positivity bias in favor of anonymous peers. *Journal of Personality and Social Psychology, 73,* 885–901. http://dx.doi.org/10.1037/0022-3514.73.5.885

Korownyk, C., Kolber, M. R., McCormack, J., Lam, V., Overbo, K., Cotton, C., . . . Allan, G. M. (2014). Televised medical talk shows—What they recommend and the evidence to support their recommendations: A prospective observational study. *British Medical Journal, 349,* g7346. http://dx.doi.org/10.1136/bmj.g7346

Kruschke, J. K. (2013). Bayesian estimation supersedes the t test. *Journal of Experimental Psychology: General, 142,* 573–603. http://dx.doi.org/10.1037/a0029146

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin, 108,* 480–498. http://dx.doi.org/10.1037/0033-2909.108.3.480

Kuzmanovic, B., & Rigoux, L. (2017). Valence-dependent belief updating: Computational validation. *Frontiers in Psychology, 8,* 1087. http://dx.doi.org/10.3389/fpsyg.2017.01087

Li, J., Delgado, M. R., & Phelps, E. A. (2011). How instructed knowledge modulates the neural systems of reward learning. *Proceedings of the National Academy of Sciences of the United States of America, 108,* 55–60. http://dx.doi.org/10.1073/pnas.1014938108

Lorenz, J., Rauhut, H., Schweitzer, F., & Helbing, D. (2011). How social influence can undermine the wisdom of crowd effect. *Proceedings of the National Academy of Sciences of the United States of America, 108,* 9020–9025. http://dx.doi.org/10.1073/pnas.1008636108

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology, 2,* 175–220. http://dx.doi.org/10.1037/1089-2680.2.2.175

Olsson, A., & Phelps, E. A. (2007). Social learning of fear. *Nature Neuroscience, 10,* 1095–1102. http://dx.doi.org/10.1038/nn1968

Ong, D. C., Zaki, J., & Goodman, N. D. (2015). Affective cognition: Exploring lay theories of emotion. *Cognition, 143,* 141–162. http://dx.doi.org/10.1016/j.cognition.2015.06.010

Oswald, M. E., & Grosjean, S. (2004). Confirmation bias. In R. Pohl (Ed.), *Cognitive illusions: A handbook on fallacies and biases in thinking, judgment and memory* (pp. 79–96). Hove, UK: Psychology Press.

Sears, D. O. (1983). The person-positivity bias. *Journal of Personality and Social Psychology, 44,* 233–250. http://dx.doi.org/10.1037/0022-3514.44.2.233

Shafto, P., Eaves, B., Navarro, D. J., & Perfors, A. (2012). Epistemic trust: Modeling children's reasoning about others' knowledge and intent. *Developmental Science, 15,* 436–447. http://dx.doi.org/10.1111/j.1467-7687.2012.01135.x

Sharot, T., Korn, C. W., & Dolan, R. J. (2011). How unrealistic optimism is maintained in the face of reality. *Nature Neuroscience, 14,* 1475–1479. http://dx.doi.org/10.1038/nn.2949

Shefrin, H. (2000). *Beyond greed and fear: Understanding behavioral finance and the psychology of investing*. New York, NY: Oxford University Press.

Shepperd, J. A., Waters, E., Weinstein, N. D., & Klein, W. M. P. (2015). A primer on unrealistic optimism. *Current Directions in Psychological Science, 24,* 232–237. http://dx.doi.org/10.1177/0963721414568341

Staudinger, M. R., & Büchel, C. (2013). How initial confirmatory experience potentiates the detrimental influence of bad advice. *NeuroImage, 76,* 125–133. http://dx.doi.org/10.1016/j.neuroimage.2013.02.074

Summerfield, C., & Egner, T. (2009). Expectation (and attention) in visual cognition. *Trends in Cognitive Sciences, 13,* 403–409. http://dx.doi.org/10.1016/j.tics.2009.06.003

Surowiecki, J. (2005). *The wisdom of crowds*. New York, NY: Knopf Doubleday Publishing Group.

Sutton, R. S., & Barto, A. G. (1998). *Introduction to reinforcement learning* (1st ed.). Cambridge, MA: MIT Press.

Tetlock, P. (2005). *Expert political judgment: How good is it? How can we know?* Princeton, NJ: Princeton University Press.

Toelch, U., Bach, D. R., & Dolan, R. J. (2014). The neural underpinnings of an optimal exploitation of social information under uncertainty. *Social Cognitive and Affective Neuroscience, 9,* 1746–1753. http://dx.doi.org/10.1093/scan/nst173

Toelch, U., Bruce, M. J., Newson, L., Richerson, P. J., & Reader, S. M. (2014). Individual consistency and flexibility in human social information use. *Proceedings of the Royal Society of London B: Biological Sciences, 281,* 20132864. https://doi.org/10.1098/rspb.2013.2864

Vanderbilt, K. E., Heyman, G. D., & Liu, D. (2014). In the absence of conflicting testimony young children trust inaccurate informants. *Developmental Science, 17,* 443–451. http://dx.doi.org/10.1111/desc.12134

Waskom, M. L., Frank, M. C., & Wagner, A. D. (2017). Adaptive engagement of cognitive control in context-dependent decision making. *Cerebral Cortex, 27,* 1270–1284. http://dx.doi.org/10.1093/cercor/bhv333

Weinstein, N. D. (1980). Unrealistic optimism about future life events. *Journal of Personality and Social Psychology, 39,* 806–820. http://dx.doi.org/10.1037/0022-3514.39.5.806

Weinstein, N. D. (1987). Unrealistic optimism about susceptibility to health problems: Conclusions from a community-wide sample. *Journal of Behavioral Medicine, 10,* 481–500. http://dx.doi.org/10.1007/BF00846146

Weinstein, N. D., & Klein, W. M. (1995). Resistance of personal risk perceptions to debiasing interventions. *Health Psychology, 14,* 132–140. http://dx.doi.org/10.1037/0278-6133.14.2.132

Willis, J., & Todorov, A. (2006). First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological Science, 17,* 592–598. http://dx.doi.org/10.1111/j.1467-9280.2006.01750.x

Worthy, D. A., & Maddox, W. T. (2014). A comparison model of reinforcement-learning and win-stay-lose-shift decision-making processes: A tribute to W. K. Estes. *Journal of Mathematical Psychology, 59,* 41–49. http://dx.doi.org/10.1016/j.jmp.2013.10.001

Yaniv, I. (2004). Receiving other people's advice: Influence and benefit. *Organizational Behavior and Human Decision Processes, 93,* 1–13. http://dx.doi.org/10.1016/j.obhdp.2003.08.002

Yaniv, I., & Kleinberger, E. (2000). Advice taking in decision making: Egocentric discounting and reputation formation. *Organizational Behavior and Human Decision Processes, 83,* 260–281. http://dx.doi.org/10.1006/obhd.2000.2909

Zarnowitz, V. (1984). The accuracy of individual and group forecasts from business outlook surveys. *Journal of Forecasting, 3,* 11–26. http://dx.doi.org/10.1002/for.3980030103

Received January 27, 2017
Revision received August 1, 2017
Accepted September 14, 2017 ■