Research Article

Psychological Science
2018, Vol. 29(2) 254–265

DOI: 10.1177/0956797617731637
www.psychologicalscience.org/PS

Good Choice, Bad Judgment: How Choice Under Uncertainty Generates Overoptimism

Jordan Tong[1], Daniel Feiler[2], and Anastasia Ivantsova[1]
[1]Wisconsin School of Business, University of Wisconsin–Madison, and [2]Tuck School of Business, Dartmouth College

Abstract

We examine a fundamental feature of choice under uncertainty: Overestimating an alternative makes one more likely to choose it. If people are naive to this structural feature, then they will tend to have erroneously inflated expectations for the alternatives they choose. In contrast to theories of motivated reasoning, this theory suggests that individuals will overestimate chosen alternatives even before they make their choice. In four studies, we found that students and managers exhibited behavior consistent with naïveté toward this relationship between estimation error and choice, leaving them overoptimistic about their chosen alternatives. This overoptimism from choosing positive error is exacerbated when the true values of the alternatives are close together, when there is more uncertainty about the values of alternatives, and when there are many alternatives to choose from. Our results illustrate how readily overoptimism emerges as a result of statistical naïveté, even in the absence of a desire to justify one's decision after the choice.

Keywords

judgment under uncertainty, choice, naive intuitive statistician, overconfidence, open data, preregistered

Received 2/13/17; Revision accepted 8/22/17

Why do new projects that people choose to pursue rarely live up to their expectations? Why does the checkout line you pick seem to slow down once you join it? Why are people more often pleasantly surprised when someone else chooses for them than when they choose for themselves? We propose that new insight into questions such as these lies in the structure of choice under uncertainty and its consequences for judgment.

A key feature of a good judgment is that it is unbiased: It does not systematically err too high or too low. A key feature of a good choice is that you pick the alternative you expect to be the best (Hogarth, 2015). Clearly, unbiased judgments help you choose the best alternative. But do good choices help you make unbiased judgments?

We argue that the answer is "no": Good choices are, in fact, an obstacle for good judgment. Although this claim may initially seem counterintuitive, consider the following logic.

1. Define good choice as choosing the alternative from a set that one believes to be the most favorable, given all available information.
2. Because of uncertainty, assume that one's belief about the favorability of an alternative will randomly err high or low to varying extents.
3. When one's belief about the favorability of an alternative randomly errs high, one is more likely to choose that alternative. Conversely, if one's belief about the favorability of an alternative randomly errs low, one will be less likely to choose that alternative.

Therefore, although good choice, as defined above, leads to good alternatives being chosen more often than

Corresponding Author:
Jordan Tong, Wisconsin School of Business, University of Wisconsin–Madison, 975 University Ave., 4293 Grainger Hall, Madison, WI 53706
E-mail: jordan.tong@wisc.edu

bad alternatives, it also leads to overestimated alternatives being systematically chosen more often than underestimated alternatives. In fact, researchers have mathematically modeled this relationship between uncertainty and optimization, and how it can generate statistical bias (Harrison & March, 1984; Smith & Winkler, 2006; see also Van den Steen, 2004). If people fail to account for this structural feature of choice that gravitates them toward overestimated alternatives, then they will be predictably overoptimistic in their chosen alternatives. According to this theory and perhaps counterintuitively, they will overestimate the chosen alternative even before having made the choice. In other words, good choice operates as a selection process—from a pool of unbiased beliefs about alternatives, it selects a subset for which one has biased judgments.

Traditional psychological research has often deemphasized structural and ecological determinants of behavior, neglecting their causal and temporal priority over motivational, emotional, and cognitive factors (Fiedler & Kutzner, 2015). However, a more recent movement of ecological psychologists have drawn inspiration from Brunswik's (1956) now-classic work, focusing on how accounting for the structure of the environment in which decisions occur often provides a simpler and more robust explanation of behavior (Fiedler, 2000; Juslin, Winman, & Hansson, 2007). According to this literature, humans can be characterized as naive intuitive statisticians: excellent at cognitively processing given samples but poor at metacognitively adjusting for the ways in which the samples generated by the environment are misrepresentative or limited (Feiler, Tong, & Larrick, 2013; Fiedler & Juslin, 2006; Tong & Feiler, 2017). This perspective attempts to provide an integrative framework for understanding a variety of human judgment biases through a single mechanism—a failure to correct for how the environment biases the information people experience. We built from this perspective to explore how simple naïveté toward the structure of choice under uncertainty can drive overoptimism.

Psychologists have long known that overoptimism can come from a different source: motivated reasoning (Kunda, 1990). After making a choice, individuals may convince themselves that the chosen alternative is outstanding in an effort to view themselves and their decision-making prowess more favorably (Sivanathan, Molden, Galinsky, & Ku, 2008). The ecological approach we took enabled us to start with the same outcome—overoptimism in chosen alternatives—and demonstrate how it can be the product of the choice-uncertainty relationship rather than a self-serving interpretation of reality. Motivated reasoning implies that choice acts as a treatment—overestimation of the chosen alternative emerges after the choice is made. In contrast, our theory implies that choice acts as a selection mechanism—chosen alternatives were already overestimated before, or at the time of, the choice. This distinction is important because the two mechanisms have very different implications for when to expect overoptimism and how it can be mitigated. For example, our theory suggests that even the most level-headed decision maker, who has absolutely no motivation to aggrandize the alternative he or she has chosen, may still exhibit overoptimism (see also Klayman, Soll, Gonzalez-Vallejo, & Barlas, 1999; Moore & Healy, 2008).

When Good Choice Is an Obstacle for Unbiased Judgment

Clearly, if you have perfectly accurate beliefs about the true values of each alternative, then choosing the alternative you think is the best will not change that. For good choice to introduce bias, there must be uncertainty in the value of alternatives and to a sufficient extent that random error in beliefs can potentially be the pivotal determinant of choice. We explored three structural factors that moderate this likelihood (Harrison & March, 1984; Smith & Winkler, 2006; Van den Steen, 2004).

First, random error is more likely to be the pivotal determinant of choice when there is less dispersion of the true values of the alternatives (i.e., they are closer together). When the true values of all alternatives are identical, your choice is entirely determined by which alternative you most overestimated. When one alternative is extraordinarily superior to the others, you will choose it even with considerable uncertainty surrounding each alternative.

Second, random error is more likely to be the pivotal determinant of choice when there is greater uncertainty about the value of alternatives. When you face greater uncertainty, you make larger valuation errors. In turn, larger errors increase the likelihood that error determines which alternative you choose.

Third, random error is more likely to be the pivotal determinant of choice when there is a greater number of alternatives to choose from. With more alternatives, it is more likely that you will grossly overestimate at least one alternative's value and consequently choose it.

Therefore, we predicted that a chosen alternative would be more severely overestimated when (a) the true values of the set of alternatives are less dispersed, (b) people face greater uncertainty about the values of alternatives, and (c) there are more alternatives to choose from (see the Supplemental Material available online for simulations of these predicted moderators). In four studies, we tested for overoptimism in chosen

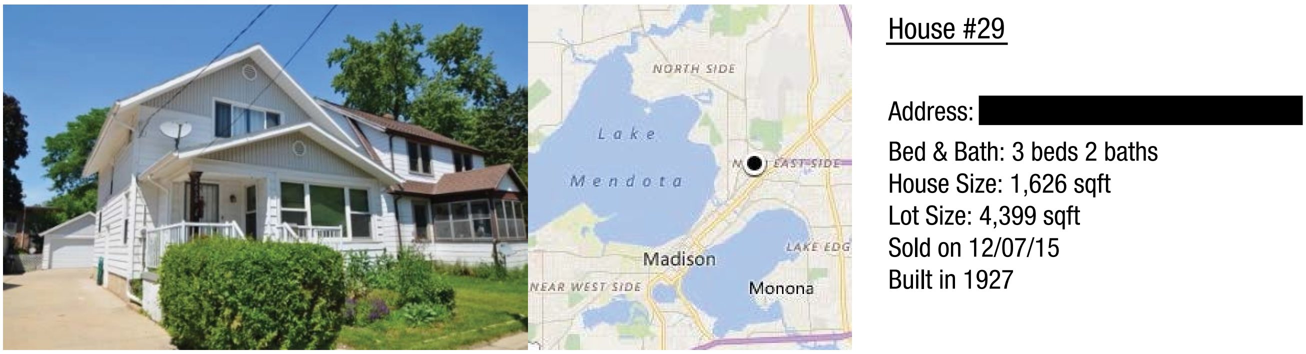

Fig. 1. Example of a house profile from Study 1. The address of the house has been redacted here for reasons of privacy.

alternatives as a result of choosing positive error. We examined whether this effect can be observed over and above motivated reasoning and whether structural factors moderated overoptimism in accordance with our theory.[1]

Study 1: Postchoice Real Estate Price Estimation

This study resembled the real-world challenge of choosing and placing valuations on uncertain assets. Subjects examined sets of real estate profiles from the local housing market, chose the house they believed to be the most valuable, and estimated its price.

Method

Subjects. We recruited subjects via an online scheduling system at a public university in the United States; subjects resided in the same city from which the real estate sample was generated. We targeted a sample size of 75 subjects by posting five laboratory sessions (the capacity was 21 per session). A total of 84 subjects participated in the study, all undergraduate (76%) or graduate (24%) students. The majority (96%) of subjects were full-time students; nearly two-thirds also had a part-time or full-time job. The sample of subjects had the following characteristics: 66% were female; 60% self-identified as White, 37% as Asian, and 2% as Black; 70% selected English as their first language; and 92% had lived in the United States for at least 1 year, while 70% had lived in the United States for at least 5 years. Each subject completed 30 rounds in total (15 in each of two conditions), which yielded a sample of 2,520 nonindependent observations. All subjects' data were included in analysis. The study was conducted in a computer lab.

Task. From Zillow.com, a leading real estate website, we collected information on single-family houses that had sold in the preceding 6 months in the local market, generating a total of 217 house profiles. For each house, we collected the following information: primary photo, address, number of bedrooms and bathrooms, size of the house and its lot (in square feet), year it was built, date it last sold, screenshot of its location on Google maps, and recent actual sale price. In the task, subjects were asked to estimate the sale price of a house, given all of the other pieces of information in the profile (an example profile is shown in Fig. 1). Before beginning, subjects were shown the entire distribution of houses that would be used in the task; this information was in the form of a histogram with sale prices on the *x*-axis. They were also shown 10 example house profiles to familiarize them with the house information that would be provided. These examples were representatively sampled from the full pool such that the mean and variance of the example houses nearly perfectly matched the mean and variance of the overall pool.

The study had a 2 × 1 within-subjects design. All subjects completed both conditions, and we randomized the order in which the conditions were completed. No house was ever shown more than once to a subject.

In each round of the *choose-estimate* condition, subjects were shown the profiles for six randomly selected houses. From this set, they first chose the house they believed sold for the highest price. Second, they estimated the chosen house's sale price. This process was repeated for 15 rounds. After each round, they received feedback regarding their performance—specifically, whether they had chosen the house that had in fact sold for the highest price and what their chosen house's actual sale price was. In each round of the *random-estimate* condition, subjects were shown only one randomly selected house at a time. Subjects simply provided an estimate of that house's sale price. This process was repeated for 15 rounds with feedback on each house's actual sale price.

These choices and estimates were incentivized. Subjects received points for the precision of their estimates and for correctly choosing the houses with the highest selling price (only in the choose-estimate condition). In the choose-estimate condition, subjects received 3 points if the house they chose each round indeed had the highest sale price in the set. In both conditions, subjects received 1, 2, or 3 points each round if their estimate was within $50,000, $25,000, or $10,000 of the true selling price, respectively. Subjects received $0.25 per point received in the game. The average payout per subject was $14.89.

Results

The dependent variable was the error of the individual's estimate in dollars, calculated as the individual's estimate of a house's price minus the actual sale price of that house. Positive errors represented overestimation of house prices, while negative errors represented underestimation.

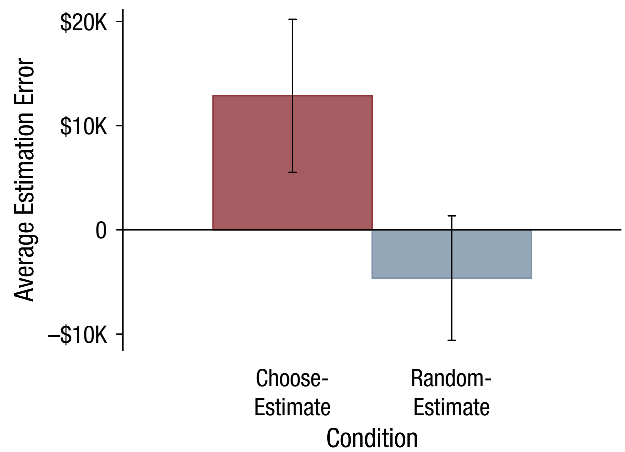

Fig. 2. Average estimation error for the choose-estimate and random-estimate conditions in Study 1. The error bars represent 95% confidence intervals. On the *y*-axis, positive values indicate overestimation of true house prices, and negative values indicate underestimation.

Simple test of overestimation of the chosen house's value. First, we examined average estimation errors by subject. Specifically, we averaged the 15 estimation errors for each subject within each condition, yielding two observations per subject: an average error in the choose-estimate condition and an average error in the random-estimate condition. We then tested whether these (a) individually differed from 0 using one-sample *t* tests and (b) differed from each other using a paired-samples *t* test.

In the choose-estimate condition, subjects significantly overestimated the actual house prices, M = $12,870, SE = $3,693, $t(83) = 3.48$, $p < .001$. However, in the random-estimate condition, in which subjects simply estimated the sale price of a single house at a time, average estimates were not significantly different from the true sale prices, M = −$4,267, SE = $3,001, $t(83) = -1.54$, $p = .13$. On average, estimates (relative to the respective true price) were $17,497 higher in the choose-estimate condition than in the random-estimate condition, SE = $4,164, $t(83) = 4.20$, $p < .001$, $d = 0.57$ (see Fig. 2). These results suggest that subjects overestimated the sale price when they chose the house from a set of six but not when a single house was randomly assigned to them.

Regression model testing overestimation of the chosen house's value. A more comprehensive analysis of the data accounted for the fact that participants more often chose to estimate the cost of expensive houses in the choose-estimate condition than in the random-estimate condition because subjects were attempting to choose the most valuable house in the former. The regression framework also enabled us to achieve additional statistical power by examining each of a subject's 30 estimates as an observation while still accounting for the within-subjects nature of the data.

We implemented a regression model with standard errors clustered by subject and fixed effects included for each house. The house fixed effects controlled for any systematic misestimation of a given house (high or low) and enabled us to, with a dummy variable for condition, answer the following question: For any given house, would we expect a different price estimate if it were chosen from a set than if it were randomly selected?

The results from this regression model show that, for a given house, subjects estimated a higher sale price if they chose the house (choose-estimate condition) than if that house was shown on its own (the random-estimate condition), β = $44,713, SE = $4,375, 95% confidence interval (CI) = [$36,011, $53,415], $p < .001$. The larger effect size relative to the simple test is due to the fact that, in general, high-priced houses tended to be underestimated, and they were chosen more frequently in the choose-estimate condition than randomly appeared in the random-estimate condition. Details of this model and its complete results can be found in the Supplemental Material.

For evidence of robustness, we also show results in the Supplemental Material from models with the following independent variables included: (a) the order in which the conditions were completed and its interaction with experimental condition, (b) the round that the estimate took place and its interaction with experimental condition, and/or (c) fixed effects included for each subject. We also re-estimated all models with an alternative dependent variable: the percentage of error,

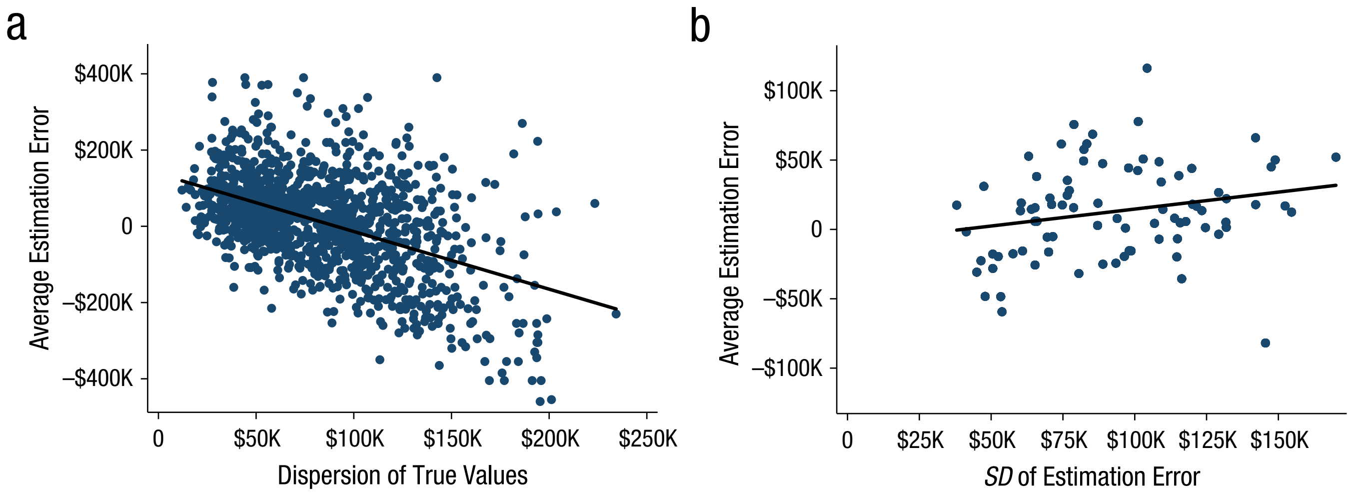

Fig. 3. Scatterplots (with trend lines) showing how overoptimism (measured by the average estimation error) in the choose-estimate condition changed with (a) the dispersion of the true values and (b) the standard deviation of estimation errors in the random-estimate condition in Study 1.

calculated as the error of the estimate divided by the true house price. Across all models, the effect of experimental condition was statistically significant, in the predicted direction, and substantial in size.

Dispersion of the true values of alternatives as a moderator. We predicted that subjects would more severely overestimate their chosen alternative's value when the true values of alternatives from which they chose were clustered closer together (i.e., less dispersed). We operationalized dispersion in true values as the standard deviation of the prices in the six-house sample observed by the subject in the choose-estimate condition.

There was a significant negative correlation between the extent of overestimation and the standard deviation of the six-house sample from which they chose, $r = -.48$, $p < .001$ (see Fig. 3a). A more comprehensive regression model, with standard errors clustered by subject and controlling for condition order and round, revealed that a $1,000 decrease in the standard deviation of the six-house sample was associated with a $1,520 increase in the degree of overestimation of one's chosen house, $\beta = -1.52$, $SE = 0.10$, 95% CI = [−1.72, −1.31], $p < .001$. This result suggests that subjects tended to overestimate the value of their chosen house to a greater extent when the true house prices from which they chose were clustered closer together.

Uncertainty about the true values of alternatives as a moderator. We predicted that subjects would more severely overestimate their chosen alternative's value when they had greater uncertainty. We operationalized uncertainty at the level of subjects by examining their degree of noise when estimating one house at a time. For each subject, we calculated the standard deviation of estimation errors in the random-estimate condition. Regarding the house prices, subjects with high standard deviations were generally more uncertain, and subjects with low standard deviations were generally less uncertain. We then used a subject's uncertainty (i.e., standard deviation of errors) during the random-estimate condition to predict his or her degree of overestimation during the choose-estimate condition.

There was a significantly positive correlation between a subject's uncertainty (in the random-estimate condition) and their average overestimation for their chosen houses (in the choose-estimate condition), $r = .23$, $p = .04$ (see Fig. 3b). A more comprehensive regression model, with standard errors clustered by subject and controlling for condition order, revealed that a $1,000 increase in a subject's uncertainty in the random-estimate condition was associated with a $300 increase in his or her degree of overestimation in the choose-estimate condition, $\beta = 0.30$, $SE = 0.14$, 95% CI = [0.01, 0.59], $p = .04$. This result suggests that subjects tended to overestimate the value of their chosen house to a greater extent when they had more uncertainty about the housing market in general.

Discussion

Study 1 provided evidence of the predicted overestimation in a task involving stimuli representatively sampled from the real world. Subjects' estimates of a single randomly selected house were not biased. However, when they chose the house that they expected to be the most valuable from a set of six, they significantly overestimated its value. For a given house, estimates were higher when it was chosen than when it was randomly assigned. Overestimation was more severe when the true values of the houses were close together and when subjects had greater uncertainty about house prices in general.

Study 2: Prechoice Real Estate Price Estimation

This study tested whether, consistent with our theory, individuals overestimate the value of chosen alternatives even before the choice. In contrast, motivated reasoning would cause belief inflation after the choice.

Method

Subjects. We recruited subjects from the same pool of college students and using the same procedures as in Study 1, resulting in a nonoverlapping sample of 86 subjects. In the sample, 83% were undergraduate students, and 17% were graduate students; 92% were full-time students, and 66% had a part-time or full-time job; 66% were female; 70% self-identified as White, 24% as Asian, and 2% as Black; 83% selected English as their first language; and 91% had lived in the United States for at least 1 year, while 86% had lived in the United States for at least 5 years.

Task. This study used the same stimuli and task as the previous study: real estate information for recently sold houses in the local market acquired from Zillow.com. The key difference was that all subjects now estimated the value of houses before even knowing that they would later need to choose the house that they believed to be worth the most. In the previous study, only the value of the chosen house was estimated, and this estimation occurred after the choice.

The task proceeded as follows. Individuals observed the same introductory information, histogram, and sample houses as in the previous study. They were then asked to estimate the sale price of 18 houses (presented in three sets of six each) given the Zillow profile information of each. Subjects received 1, 2, or 3 points each round if their estimate was within \$50,000, \$25,000, or \$10,000 of the true selling price, respectively. Subjects received \$0.25 per point earned.

Next, subjects were informed that they would be shown the three sets of six houses again. For each set of six, they needed to try to choose the house that was the most valuable—the one that had sold for the highest price. For each of the three rounds, subjects received 3 points (\$0.75) if they chose the most valuable house in the set. The average payout per subject was \$10.35.

Results

The dependent variable was the subject's estimate of a house's price minus the actual sale price of that house: the error of the estimate. There were two recorded estimations that were orders of magnitude different from the other estimations and likely the product of typographical errors. As specified in the preregistration of this study (via the Open Science Framework), we excluded the observations from these two rounds. No analyses were conducted before these observations were identified for exclusion. Post hoc analyses showed that their omission did not substantively change the results, but their exclusion provides a more reliable estimate of the true effect size.

Simple test of overestimation of the chosen house's value. To begin with a simple and direct analytical approach, we examined average estimation errors by subject. First, we computed each subject's average estimation error for houses in general (18 estimations per subject). Second, we computed each subject's average estimation error for the houses that were subsequently chosen (3 estimations per subject). We then tested whether each of these differed from 0 (via one-sample t tests) and from each other (via a paired-samples t test).

In general, the average estimates of all house prices were not significantly different from their true prices, $M = -\$5,307$, $SE = \$3,899$, $t(85) = -1.36$, $p = .18$. However, the prices of subsequently chosen houses were significantly overestimated, $M = \$37,154$, $SE = \$9,035$, $t(85) = 4.11$, $p < .001$. On average, estimates (relative to the respective true price) were \$42,461 higher for the subset of houses that were subsequently chosen than for houses in general, $SE = \$6,921$, $t(85) = 6.14$, $p < .001$, $d = 0.65$ (see Fig. 4). These results suggested that before choosing—and before even being aware that they would later be making a choice—subjects already overestimated the sale price of the house that they later chose. This pattern of results cannot be accounted for by motivated reasoning because the estimates occurred before the choice.

Regression model testing overestimation of the chosen house's value. As in Study 1, we also implemented a regression framework. A least-squares regression model was conducted with a dummy variable for whether or not the estimation was for a house that was subsequently chosen, standard errors clustered by subject, and fixed effects included for each house. The house fixed effects enabled us to examine the following question: For any given house, should we expect higher estimation errors if the subject subsequently chose that house from a set of six than if it was not subsequently chosen? The results

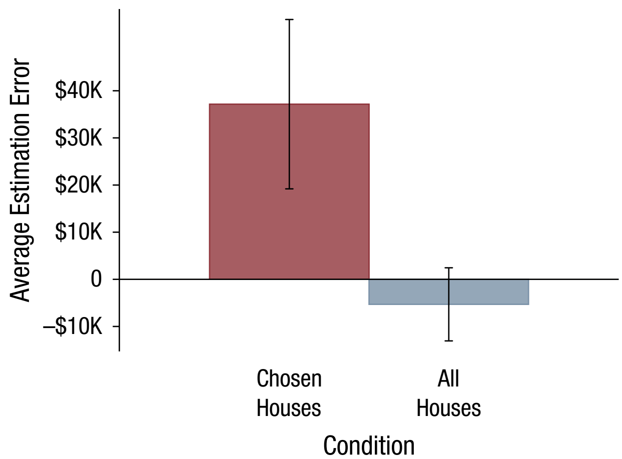

Fig. 4. Average estimation error for chosen houses and all houses in Study 2. The error bars represent 95% confidence intervals. On the *y*-axis, positive values indicate overestimation of true house prices, and negative values indicate underestimation.

from this regression model show that subjects made more positively biased estimation errors for a given house if they went on to choose it than if they did not eventually choose it, β = $71,832, *SE* = $9,354, 95% CI = [$53,233, $90,431], $p < .001$.

For evidence of robustness, we also show results in the Supplemental Material from models with the following independent variables included: (a) a dummy variable for whether or not the estimation was for a house that was subsequently chosen, (b) the round that the estimate took place, (c) the interaction between the two, and/or (d) fixed effects for each subject. We also re-estimated these models with the dependent variable specified as percentage of error. Across all of these models, having subsequently being chosen was significantly predictive of a house having been overestimated in the first place. Complete details and results for these models are presented in the Supplemental Material.

Finally, consistent with Study 1, results showed support for two moderating factors. There was a significant negative correlation between the extent of overestimation for chosen houses and the standard deviation of the six-house sample from which subjects chose, $r = -.34$, $p < .001$. Also, there was a significantly positive correlation between subjects' uncertainty about the true values of alternatives in general (operationalized as the standard deviation of errors for all house estimates) and their average overestimation of their chosen houses, $r = .24$, $p = .029$. Further regression models on these moderators are available in the Supplemental Material.

Discussion

In this study, subjects estimated housing prices before they knew they would later make a choice. Subjects' estimates were unbiased in general; however, the houses subsequently chosen were already overestimated in prechoice estimates. This result is consistent with our theory that choice acts as a selection mechanism; it cannot be explained by a theory of choice acting as a treatment mechanism.

Study 3: Hiring and Ability Estimation

Study 3 was a simpler, one-shot task with a Bayesian solution. The task was a hiring decision informed only by a single numerical signal of each job candidate's ability.

Method

Subjects. We targeted a sample size of 400 subjects. The final sample consisted of 489 managers (40.2% female, mean age = 47.1 years) who were reached via ROI Rocket, a survey company that maintains a set of professionals and consumers who have indicated in the past that they are interested in completing surveys in exchange for compensation. These individuals were invited via e-mail to participate in a "Management Survey." The survey was closed 2 days after the invitation to participate. No analyses were run on any preliminary subset of the data. Subjects who took less than 3 min or more than 45 min were excluded from analyses on the basis of predetermined criteria for what was a reasonable amount of time to complete the task (19 observations were removed). The subjects were in managerial positions in the United States, were between the ages of 25 and 60 years, had earned at least a bachelor's degree, and had at least one subordinate reporting to them at their place of work. Subjects received between $4 and $5.50 in exchange for their participation in a 5-min survey.

Procedure and experimental design. Subjects played the role of a chief recruiter at a management consulting firm. They decided which hypothetical candidate to hire for an entry-level position. Each job candidate had taken a test that served as a noisy measure of their problem-solving ability. Problem-solving ability was the key attribute that determined worker productivity in the firm; therefore, subjects were attempting to hire the candidate with the highest problem-solving ability.

We administered an online survey programmed in Qualtrics. Subjects knew (a) the distribution of true abilities across candidates and (b) the distribution of test-score measurement error. Subjects were told the following: The population of candidates' true abilities was normally distributed with mean μ_{ability} of 100 and the standard deviation σ_{ability} of 10, which was also depicted in a histogram. However, each individual candidate's true ability could not be observed directly.

Instead, all candidates had taken a problem-solving test. Each candidate's test score was observable and was a noisy measure of their true problem-solving ability. As a measure of true ability, the test score had normally distributed measurement error with a mean of 0 and standard deviation $\sigma_{\text{test\_score}}$ of 25, which was also depicted in a histogram.

Subjects were shown a set of candidates (which had been randomly selected from the distribution described above) along with each candidate's test score (randomly generated according to the measurement error described above). First, subjects chose which candidate they wanted to hire. If they selected the candidate with the highest true problem-solving ability in the set, then they earned a \$1 bonus. On the subsequent screen, they were asked to estimate the true problem-solving ability of the candidate they had hired. They earned an accuracy bonus for this estimate: \$0.50 minus their absolute error (the distance between their estimate and the true ability of the candidate they selected), with a minimum of \$0. Subjects were randomly assigned to two experimental conditions, which differed only in the number of candidates from which they could choose: 3 (the *3-alternative* condition) or 10 (the *10-alternative* condition). Otherwise, the conditions were identical.

Bayesian solution. In this study, there was a clearly defined statistical solution to the problem that properly corrected for the fact that the best test score tended to have benefitted from positive error (for details, see Smith & Winkler, 2006). As adapted for our study, let μ_{ability} be the population's average true problem-solving ability. Let $\sigma^2_{\text{ability}}$ be the variance of the population's true problem-solving abilities. Let $\sigma^2_{\text{test\_score}}$ be the variance of the test's measurement error. Then, given a candidate's test score x, the Bayesian estimate for that candidate's true intelligence is $e^{\text{Bayes}} = \alpha x + (1 - \alpha)\ \mu_{\text{ability}}$, where $\alpha = 1/(1 + \sigma^2_{\text{test\_score}}/\sigma^2_{\text{ability}})$. Thus, the best estimate of a candidate's true ability is a weighted average of (a) the candidate's test score and (b) the average ability in the population. The smaller the measurement error of the test, the more weight one should give the test score relative to the population's average ability. The less variance in ability there is in the population, the more weight one should give the population's average ability. Here, $\mu_{\text{ability}} = 100$, and

$$\alpha = \frac{1}{1 + \frac{25^2}{10^2}} \approx 13.8\%, \text{ so } e^{\text{Bayes}} \approx .138x + 86.2.$$

Therefore, in this study, a "perfectly rational" automated Bayesian player would choose the candidate with the highest test score (or equivalently, the candidate with the highest e^{Bayes} value) and then estimate that candidate's ability to be e^{Bayes}. We used e^{Bayes} as a benchmark for comparison with the estimates of subjects.

Results

Estimates of true ability. The dependent variable was the estimation error: the subject's estimate of the chosen candidate's true ability minus that candidate's actual true ability. Overall, there was a significant tendency to overestimate the ability of the candidate that one had chosen to hire, $t(488) = 7.51$, $p < .001$. In the 3-alternative condition, on average, subjects overestimated the hired candidate's ability by 1.86, $SE = 1.09$, $t(255) = 1.71$, $p = .09$. In the 10-alternative condition, on average, subjects overestimated the hired candidate's ability by 11.22, $SE = 1.22$, $t(232) = 9.18$, $p < .001$, $d = 1.12$. There was significantly greater overestimation of the ability of the hired candidate in the 10-alternative condition than in the 3-alternative condition, $t(487) = 5.73$, $p < .001$, $d = 0.52$.

In contrast, an automated Bayesian player did not display overoptimism. Given the identical set of scenarios faced by subjects, the Bayesian player's average error was only −0.05 ($SE = 0.58$) in the 3-alternative condition and 0.23 ($SE = 0.64$) in the 10-alternative condition; neither was significantly different from zero, $t(255) = -0.09$, $p = .93$, and $t(232) = -0.37$, $p = .71$, respectively (see Fig. 5).

Test scores of chosen candidates. To explore how this overestimation emerged, we examined the test scores of the candidates hired by subjects. Overall, the test score of the chosen candidate was biased higher than the candidate's true ability, $t(488) = 13.05$, $p < .001$. In the 3-alternative condition, on average, the test score of the hired candidate was 7.30 ($SE = 1.22$) higher than his or her true ability, $t(255) = 5.95$, $p < .001$. In the 10-alternative condition, on average, the test score of the hired candidate was 18.86 ($SE = 1.46$) higher than his or her true ability, $t(232) = 12.91$, $p < .001$. Thus, the test scores of the chosen candidates were biased high, even though test scores in general were unbiased.

Subjects did appear to adjust their guesses downward from the observed test scores toward the population mean of 100. They adjusted downward by 5.44 in the 3-alternative condition, $t(255) = -4.91$, $p < .001$, and by 7.64 in the 10-alternative condition, $t(232) = -6.80$, $p < .001$. However, they did so insufficiently and less than a Bayesian player would, which is why we observed significant overestimation of ability. Although subjects adjusted slightly more in the 10-alternative condition than in the 3-alternative condition, the difference in adjustment magnitude was not significant, $t(487) = 1.39$, $p = .16$.

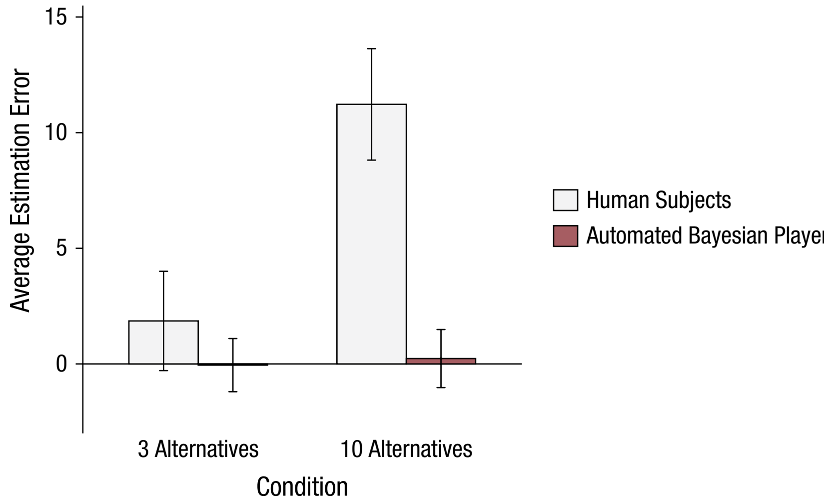

Fig. 5. Average estimation error in the 3- and 10-alternative conditions in Study 3, separately for human subjects and an automated Bayesian player. The error bars represent 95% confidence intervals.

Discussion

In a simple setting with a Bayesian solution, subjects still overestimated chosen alternatives. People failed to make sufficient statistical adjustments, even though the task required no interpretation of ambiguous information—the signals for each alternative were objectively provided. This overestimation was worse when subjects were given more alternatives from which to choose.

Study 4: Initial Estimates and Independent Judgments With Jars

This study replicated the overestimation result of the previous studies and explored three additional questions. Are prechoice and postchoice estimates consistent? Can one reduce overestimation by having a different person estimate the alternative selected by the chooser? Can the effect be reflected such that individuals also underestimate the value of the alternative they think is the worst?

Method

Subjects. We implemented the following recruiting strategy to implement a paired-subjects 2 × 1 design. For the *choice-estimate* condition, we targeted 50 subjects by recruiting through an online scheduling system using a predetermined 50 sessions (capacity = 1 subject per session), which yielded 44 subjects. We then recruited subjects for the *independent-estimate* condition using the same online scheduling system and matched subjects sequentially to the subjects in the first condition until we obtained 44 subjects for that condition as well, which yielded 88 total subjects (59% female). Subjects were recruited from the same pool as in Study 1, although no subjects participated in both studies.

Design and procedure. Subjects were told that they would play the role of a business analyst who values projects. There were six projects, each of which was represented by a unique and nontransparent jar filled with an unknown number of pennies. The study had a 2 × 1 paired-samples between-subjects design. Subjects in the choice-estimate condition ("choosers") were matched with subjects in the independent-estimate condition ("nonchoosers") such that both made incentivized estimates for the exact same jars.

For choosers, there were five steps. First, they were asked to physically examine the six jars and estimate the amount of money contained in each jar. Second, they were asked to choose the jar they believed to contain the most money (the "chosen maximum jar"), earning $2 if they chose correctly. Third, they were asked to re-estimate the amount of money contained in the jar they had chosen as the most valuable. Fourth, subjects were asked to choose the jar they believed to contain the least money (the "chosen minimum jar"), earning $2 if they chose correctly. Finally, subjects were asked to re-estimate the amount of money contained in the jar they had designated as the least valuable. In both final estimations, subjects earned a bonus of $2 for an estimate within $0.10 of the true value of that jar and $1 for an estimate within $0.50. During the initial estimations (Step 1), choosers were unaware that they would next be choosing a maximum and minimum jar.

Nonchoosers were each linked with a unique chooser who had just completed the study; the nonchoosers were blind to the chooser's identity. The nonchooser was shown only the two jars that had been chosen by the linked chooser (the chosen maximum jar and the chosen minimum jar) and asked to estimate the amount of money in each. Incentives were the same for the nonchoosers. Nonchoosers were not told that the two jars had been chosen as the maximum and minimum by another subject.

Results

Chosen maximum jar. In their final estimates—which occurred after the choice—choosers significantly overestimated the jar they had chosen as containing the most money, M = \$2.76, SE = \$0.32, $t(43) = 8.58$, $p < 0.001$. Importantly, in Step 1, subjects already significantly overestimated the value of the jar that they later went on to choose as the most valuable, M = \$2.41, SE = \$0.35, $t(43) = 6.80$, $p < .001$. There was no significant change in beliefs between initial and final estimates, $t(43) = 1.63$, $p = .11$, suggesting that the ultimate overestimation of chosen jars was a product of initial errors in beliefs. Nonchoosers did not overestimate the value of their respective chooser's chosen maximum jar. On average, nonchoosers' final estimates were not significantly different from the true value, M = \$0.51, SE = \$0.32, $t(43) = 1.58$, $p = .12$. A pairwise comparison revealed that choosers' estimates were on average \$2.26 higher than their linked nonchoosers', $t(43) = 4.63$, $p < .001$, $d = 1.06$.

Chosen minimum jar. A similar, but reflected, pattern of results was observed for the jar that the chooser had selected as the least valuable. Choosers significantly underestimated the chosen minimum jar in their final estimates, M = –\$0.62, SE = \$0.13, $t(43) = -4.74$, $p < .001$. The initial estimate of the specific jar that the subject would later choose as the least valuable was already biased low, M = –\$0.63, SE = \$0.16, $t(43) = -3.98$, $p < .001$. There was no significant change in beliefs between initial and final estimates, $t(43) = 0.035$, $p = .972$. On average, nonchoosers' estimates of the chosen minimum jars were not significantly different from their true values, M = \$0.20, SE = \$0.19, $t(43) = -1.05$, $p = .30$. A pairwise comparison revealed that for the chosen minimum jars, choosers' estimates were on average \$0.42 lower than their linked nonchoosers', SE = \$0.23, $t(43) = -1.82$, $p = .08$, $d = -0.38$.

Discussion

After trying to choose the most valuable jar, subjects systematically overestimated its value. Before the choice, subjects already overestimated the value of the jars they went on to choose, and they did not significantly change these beliefs after choosing. Independent judges who did not make a choice were unbiased in their estimates of the jars chosen by other subjects, consistent with the idea that people's estimation errors are idiosyncratic. Lastly, the effect was reflected: Subjects underestimated the value of jars they chose as least valuable.

General Discussion

Four studies found that people make overoptimistic judgments of chosen alternatives, even with unbiased judgments in general. Consistent with the theory that overoptimism can be due to naïveté toward the structural relationship between error and choice, individuals tended to already overestimate whichever alternative they ultimately chose even before making the choice. Also consistent with this theory, the bias was worse when the true values of the alternatives were closer together, when the estimation errors were larger in general, and when there were more alternatives to choose from. This overoptimism persisted even when there was an objective mathematical solution. The effect was reflected when individuals were asked to judge the value of the alternative chosen as the worst. It was reduced when the estimation of the chosen alternative's value was done by a different person than the chooser.

Where traditional psychology has largely focused on factors within individuals to understand behavior, a recent movement in psychology stresses the importance of studying the environment and how it shapes behavior (Fiedler, 2000; Juslin et al., 2007). This perspective helps integrate many behavioral phenomena under a single mechanism: a failure to correct for the misrepresentatives samples people experience in their environments. We contribute to this stream by focusing on how the fundamental structure of choice under uncertainty becomes an obstacle for good judgment.

Our findings also contribute to the literature on overoptimism. Considerable research has found that people often engage in wishful thinking (Armor, Massey, & Sackett, 2008; Krizan & Windschitl, 2007). They believe an outcome is more likely when they view it as more desirable, even with experience and incentives for accuracy (Massey, Simmons, & Armor, 2011; Simmons & Massey, 2012). Many existing psychological theories can explain why one might change one's belief about a chosen alternative after choice, because of either a desire to view oneself favorably (Kunda, 1990; Leary & Kowalski, 1990) or self-perception dynamics (Bem, 1967; Festinger, 1957). Where previous work has focused on motivational, emotional, "hot" processing explanations for overoptimism, we have studied a dispassionate, cognitive, "cool" processing explanation for overoptimism. In this way, our findings complement the literature on cool processing explanations for various biases, such as overconfidence as a product of noisy beliefs (Hogarth & Karelaia, 2012; Klayman et al., 1999; Moore & Healy, 2008; Moore & Small, 2007), nonregressive thinking when predicting events in sequence (Kahneman & Tversky, 1973), and overbidding in auctions (Bazerman & Samuelson, 1983).

From an applied perspective, our results emphasize how easy it is to become overoptimistic about the courses of action one selects and the ventures one chooses to pursue. Funding agencies, policymakers, managers, and individuals often must choose a project to support or avenue to pursue from a set of alternatives with uncertain benefits. Our results provide a warning that even unselfish, level-headed people are likely to be overoptimistic about what they choose if they are naive to the relationship between uncertainty and choice.

Action Editor

Timothy J. Pleskac served as action editor for this article.

Author Contributions

J. Tong and D. Feiler jointly developed the theoretical concept. Data collection was conducted by A. Ivantsova. All authors jointly designed the studies, analyzed the data, and drafted the manuscript. All authors approved the final version of the manuscript for submission.

Acknowledgments

We thank Justin Sydnor, Eesha Sharma, Don Hausch, Gregory DeCroix, Jim Morris, Enno Siemsen, Bob Batt, Hessam Bavafa, Ron Adner, Syd Finkelstein, and Andy King for their insightful conversations and feedback.

Declaration of Conflicting Interests

The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

Supplemental Material

Additional supporting information can be found at http://journals.sagepub.com/doi/suppl/10.1177/0956797617731637

Open Practices

All data have been made publicly available via the Open Science Framework and can be accessed at https://osf.io/zg427/. The design and analysis plans for Study 2 was preregistered at the Open Science Framework (https://osf.io/q2g8n/). The complete Open Practices Disclosure for this article can be found at http://journals.sagepub.com/doi/suppl/10.1177/0956797617731637. This article has received badges for Open Data and Preregistration. More information about the Open Practices badges can be found at http://www.psychologicalscience.org/publications/badges.

Note

1. Materials, stimuli, and data for all studies are available in the Supplemental Material.

References

Armor, D. A., Massey, C., & Sackett, A. M. (2008). Prescribed optimism: Is it right to be wrong about the future? *Psychological Science*, *19*, 329–331.

Bazerman, M. H., & Samuelson, W. F. (1983). I won the auction but don't want the prize. *The Journal of Conflict Resolution*, *27*, 618–634.

Bem, D. J. (1967). Self-perception: An alternative interpretation of cognitive dissonance phenomena. *Psychological Review*, *74*, 183–200.

Brunswik, E. (1956). *Perception and the representative design of experiments*. Berkeley: University of California Press.

Feiler, D. C., Tong, J. D., & Larrick, R. P. (2013). Biased judgment in censored environments. *Management Science*, *59*, 573–591.

Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford, CA: Stanford University Press.

Fiedler, K. (2000). Beware of samples! A cognitive-ecological sampling approach to judgment biases. *Psychological Review*, *107*, 659–676.

Fiedler, K., & Juslin, P. (2006). *Information sampling and adaptive cognition*. Cambridge, England: Cambridge University Press.

Fiedler, K., & Kutzner, F. (2015). Information sampling and reasoning biases: Implications for research in judgment and decision making. In G. Keren & G. Wu (Eds.), *The Wiley Blackwell handbook of judgment and decision making* (pp. 380–403). Chichester, England: John Wiley & Sons.

Harrison, J. R., & March, J. G. (1984). Decision making and postdecision surprises. *Administrative Science Quarterly*, *29*, 26–42.

Hogarth, R. M. (2015). What's a "good" decision? Issues in assessing procedural and ecological quality. In G. Keren & G. Wu (Eds.), *The Wiley Blackwell handbook of judgment and decision making* (pp. 952–972). Chichester, England: John Wiley & Sons.

Hogarth, R. M., & Karelaia, N. (2012). Entrepreneurial success and failure: Confidence and fallible judgment. *Organization Science*, *23*, 1733–1747.

Juslin, P., Winman, A., & Hansson, P. (2007). The naïve intuitive statistician: A naïve sampling model of intuitive confidence intervals. *Psychological Review*, *114*, 678–703.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, *80*, 237–251.

Klayman, J., Soll, J. B., Gonzalez-Vallejo, C., & Barlas, S. (1999). Overconfidence: It depends on how, what, and whom you ask. *Organizational Behavior and Human Decision Processes*, *79*, 216–247.

Krizan, Z., & Windschitl, P. D. (2007). The influence of outcome desirability on optimism. *Psychological Bulletin*, *133*, 95–121.

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, *108*, 480–498.

Leary, M. R., & Kowalski, R. M. (1990). Impression management: A literature review and two-component model. *Psychological Bulletin*, *107*, 34–47.

Massey, C., Simmons, J. P., & Armor, D. A. (2011). Hope over experience: Desirability and the persistence of optimism. *Psychological Science*, *22*, 274–281.

Moore, D. A., & Healy, P. J. (2008). The trouble with overconfidence. *Psychological Review*, *115*, 502–517.

Moore, D. A., & Small, D. A. (2007). Error and bias in comparative judgment: On being both better and worse than we think we are. *Journal of Personality and Social Psychology*, *92*, 972–989.

Simmons, J. P., & Massey, C. (2012). Is optimism real? *Journal of Experimental Psychology: General*, *141*, 630–634.

Sivanathan, N., Molden, D. C., Galinsky, A. D., & Ku, G. (2008). The promise and peril of self-affirmation in de-escalation of commitment. *Organizational Behavior and Human Decision Processes*, *107*, 1–14.

Smith, J. E., & Winkler, R. L. (2006). The optimizer's curse: Skepticism and postdecision surprise in decision analysis. *Management Science*, *52*, 311–322.

Tong, J., & Feiler, D. (2017). A behavioral model of forecasting: Naive statistics on mental samples. *Management Science*, *63*, 3609–3627. doi:10.1287/mnsc.2016.2537

Van den Steen, E. J. (2004). Rational overoptimism (and other biases). *American Economic Review*, *94*, 1141–1151.