Organizational Behavior and Human Decision Processes 162 (2021) 155–188

Contents lists available at ScienceDirect

Organizational Behavior and Human Decision Processes

journal homepage: www.elsevier.com/locate/obhdp

Decision making can be improved through observational learning☆

Haewon Yoon [a,*], Irene Scopelliti [b], Carey K. Morewedge [c]

[a] *Kelley School of Business, Indiana University, United States*
[b] *The Business School (formerly Cass), City, University of London, United Kingdom*
[c] *Questrom School of Business, Boston University, United States*

ARTICLE INFO

Action Editor: M. Koutras

Keywords:
Debiasing
Social learning
Cognitive bias
Weight on advice
Knowledge transfer
Tacit knowledge

ABSTRACT

Observational learning can debias judgment and decision making. One-shot observational learning-based training interventions (akin to "hot seating") can produce reductions in cognitive biases in the laboratory (i.e., anchoring, representativeness, and social projection), and successfully teach a decision rule that increases advice taking in a weight on advice paradigm (i.e., the averaging principle). These interventions improve judgment, rule learning, and advice taking more than practice. We find observational learning-based interventions can be as effective as information-based interventions. Their effects are additive for advice taking, and for accuracy when advice is algorithmically optimized. As found in the organizational learning literature, explicit knowledge transferred through information appears to reduce the stickiness of tacit knowledge transferred through observational learning. Moreover, observational learning appears to be a unique debiasing training strategy, an addition to the four proposed by Fischhoff (1982). We also report new scales measuring individual differences in anchoring, representativeness heuristics, and social projection.

1. Introduction

People exhibit biased judgments and decisions in a host of personal and professional domains ranging from business to education, law, medicine, and public policy. Novices and experts are often similarly susceptible to these systematic deviations from the prescriptions of objective standards including facts, statistics, and logic. Realtors are just as anchored as undergraduates by manipulated listing prices in their assessments of the value of homes (Northcraft & Neale, 1987). Physicians are as influenced by framing effects as their patients when deciding between surgical and therapeutic interventions (McNeil, Pauker, Sox, & Tversky, 1982; Tversky & Kahneman, 1981). Philosophers are as susceptible as their students to decision frames when resolving moral dilemmas such as the trolley problem (Schwitzgebel & Cushman, 2015).

Not all people, however, are similarly affected by bias in their judgments and decisions. Substantial individual differences exist in the degree to which people exhibit numerous cognitive biases, including bias blind spot, correspondence bias, overconfidence, and loss aversion (Cokely et al., 2018; De Bruin, Parker, & Fischhoff, 2007; Frederick, 2005; Mellers et al., 2015; Scopelliti et al., 2015; Scopelliti, Min, McCormick, Kassam, & Morewedge, 2018; Tom, Fox, Trepel, & Poldrack, 2007). The existence of this individual level variation, coupled with the ability to debias judgment and decision making through training interventions, provides encouraging evidence that reasoning can be improved. Indeed, debiasing training interventions have reduced the influence of cognitive biases such as susceptibility to anchoring, bias blind spot, correspondence bias, overconfidence, social projection, sunk cost fallacies, and overreliance on forms of the representativeness heuristic. The resulting improved judgments and decisions, in both laboratory and field contexts, have been observed as long as three months post-intervention (Chang, Chen, Mellers, & Tetlock, 2016; Fong, Krantz, & Nisbett, 1986; Fong & Nisbett, 1991; Larrick, Morgan, & Nisbett, 1990; Morewedge et al., 2015; Nisbett, 1993; Sellier, Scopelliti, & Morewedge, 2019).

Debiasing training interventions typically incorporate some combination of four basic strategies proposed by Fischhoff (1982): (i) warning people about the possibility of bias, (ii) describing the direction in which bias may influence judgment, (iii) providing feedback on judgments and

☆ This work was supported by the Intelligence Advanced Research Projects Activity via the Air Force Research Laboratory contract number FA8650-11-C-7175 to C. K. Morewedge and a Cass Pump Priming Grant to I. Scopelliti. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of IARPA, AFRL, or the U.S. Government.

* Corresponding author at: Kelley School of Business Indianapolis, 801 W. Michigan Street, Indianapolis, IN 46202, United States.
E-mail address: haeyoon@iu.edu (H. Yoon).

https://doi.org/10.1016/j.obhdp.2020.10.011
Received 2 February 2018; Received in revised form 24 September 2020; Accepted 19 October 2020

decisions, or (iv) providing training with extensive coaching (for a review, see Soll, Milkman, & Payne, 2016). While purely information-based training interventions (i.e., i & ii) can be effective (Larrick et al., 1990; Morewedge et al., 2015), the most effective interventions appear to incorporate personalized feedback and more extensive forms of coaching and training, such as through statistics courses or "serious" video games (Fischhoff, 1982; Morewedge et al., 2015; Nisbett, Fong, Lehman, & Cheng, 1987).

We examine the efficacy of a fifth training strategy—observational learning. Observational learning is a form of social learning whereby people acquire attitudes, norms, and skills by observing and modeling the behavior of other agents (Bandura, 1977; Nadler, Thompson, & Boven, 2003; Tim & Luthans, 1980). It allows people to learn by observing rather than experiencing the direct consequences of a behavior. Social animals learn risks and rewards, foundational attributes underlying decision making, by observing the consequences of others' decisions and actions (Danchin, Giraldeau, Valone, & Wagner, 2004; Frith & Frith, 2012; Kline, 2015). Monkeys learn the location of food, dangerous objects, action sequences, and social ranks, by observing the behavior of conspecifics (Frith & Frith, 2012). Human infants make inferences about objects (e.g., "I can eat that!") and actors (e.g., "I can't trust him.") by observing other agents (Choi & Luo, 2015; Gergely, Bekkering, & Király, 2002; Kuhlmeier, Wynn, & Bloom, 2003; Wertz & Wynn, 2014). In organizational contexts, employees learn norms and behaviors by watching the behavior and interactions of their coworkers and managers (for a review, see Tim & Luthans, 1980), and managers can train their employees by directly modeling desired behaviors (Manz & Sims, 1981).

Observational learning is a route through which cognitive biases are inculcated and exacerbated. Judgmental bias can be amplified by observing fellow group members (Norton, Cooper, Monin, & Hogg, 2003), and through exposure to social media content shared by others that confirms people's polarized and stereotyped beliefs (Del Vicario et al., 2016; Del Vicario, Zollo, Caldarelli, Scala, & Quattrociocchi, 2017; DellaVigna & Kaplan, 2008; Mocanu, Rossi, Zhang, Karsai, & Quattrociocchi, 2015). Observational learning leads people to conform to the biased behavior of other people due to automatic and deliberate forms of social influence (Cialdini & Goldstein, 2004; Huh, Vosgerau, & Morewedge, 2014), and through other social learning mechanisms such as social signaling and culture (e.g., Bollinger & Gillingham, 2012; Maddux et al., 2010; Miyamoto & Kitayama, 2002).

We suggest that observational learning may also be a means by which to improve judgment and decision making. As children can learn abstract rules by modeling the behavior of others (Zimmerman & Rosenthal, 1974), adults can learn and improve their reasoning by observing biased and unbiased judgments and decisions made by other people. Analogous effects have been observed with respect to individuals learning motor skills and organizations learning knowledge-based skills. Seeing a person practicing a physical activity allows observers to develop mechanisms for detecting and correcting errors similar to those acquired by directly practicing the activity (Blandin & Proteau, 2000; Kappes & Morewedge, 2016). Much organizational learning happens by observing and imitating product and knowledge-based innovations developed by other organizations (Kogut & Zander, 1992; McEvily & Chakravarthy, 2002; Zander & Kogut, 1995).

The efficacy of observational learning as a debiasing training intervention, a strategy to reduce biased judgment and decision making, has not been examined. Research to date has focused on debiasing interventions in the form of incentives, nudges, and direct training, which do not require decision makers to observe the behavior of others. Even social influence-based interventions (e.g., letters reporting average energy consumption, door hangers conveying rates of towel reuse; Goldstein, Cialdini, & Griskevicius, 2008; Schultz, Nolan, Cialdini, Goldstein, & Griskevicius, 2007), typically provide people with information about the behavior of others. They do not directly expose them to the behavior of their peers (cf., Huh et al., 2014; Kraft-Todd, Bollinger, Gillingham, Lamp, & Rand, 2018). Such exposure could have increased the efficacy of these interventions. Observation is likely to engage cognitive processes similar to those occurring while performing the actual behavior (e.g., increase accessibility of relevant information, evoke inferences about debiasing strategies, associate necessary action units; Kappes & Morewedge, 2016).

We examined whether observational learning training interventions can improve judgment and decision making in three experiments. Given the general efficacy of learning by observing others model a behavior (Zimmerman & Rosenthal, 1974), particularly in related contexts such as negotiation (Nadler et al., 2003), we expected observational learning to be significantly more effective at debiasing than practice alone (i.e., making particular judgments and decisions without any training intervention; our control condition). We also predicted observational learning-based interventions would have additive training effects when paired with information-based interventions. Observational learning should provide unique benefits, in addition to information-based training.

In Experiment 1, we tested an intensive 90-minute training intervention targeting three cognitive biases (anchoring, social projection, and overreliance on the representativeness heuristic) with a variety of mitigating strategies. The experiment served as an existence proof of the efficacy of observational-learning based interventions, and it compared their efficacy relative to directly experiencing the training intervention and receiving an information-based intervention. In Experiments 2A and 2B, we focused more narrowly on teaching, in a few minutes, a single decision rule that would increase advice taking in a weight on advice paradigm. Their narrower focus allowed us to precisely examine the unique effects of observational learning. Furthermore, by varying the quality of the advice provided in these two experiments—from advice that was naturalistic (2A) to advice that was optimized by an algorithm to produce a more accurate answer if participants used the decision rule taught in the interventions (2B)—we explored when observational-learning based interventions provide additional benefits to learning, and might improve judgmental accuracy beyond comparable information-based interventions.

2. Experiment 1

Experiment 1 served as an existence proof for the efficacy of observational learning as a debiasing training intervention. In a laboratory experiment lasting approximately 3 h, we compared the efficacy of an observational learning-based training intervention to three alternative benchmarks: a control condition with no intervention and two debiasing interventions found to be effective in previous research. One debiasing intervention entailed watching an instructional video. The other entailed playing a serious video game (Morewedge et al., 2015; Sellier, Scopelliti, & Morewedge, 2019). In the observational learning intervention, participants viewed the gameplay of a participant who played the serious game. They saw this in real-time on their computer monitor, but could not otherwise see or interact with the game player.

The interventions were administered between the completion of pretest and posttest measures of three cognitive biases influential in intelligence analysis (Intelligence Advanced Research Projects Activity, 2011): anchoring, social projection, and overreliance on the representativeness heuristic. This design allowed us to compare the debiasing effects of the observational learning intervention to the three other conditions.

Our measures of anchoring and social projection (Appendices A and B, respectively) examined the effect of the interventions on increasing the correction of judgments from an initial numerical value or egocentric perspective, respectively (Epley, Morewedge, & Keysar, 2004; Robbins & Krueger, 2005; Simmons, LeBoeuf, & Nelson, 2010; Tversky & Kahneman, 1974). The measures of representativeness (Appendix C) examined the debiasing effects of the interventions on a family of biases that stem from the overgeneralization of associative similarity heuristics to cases where they violate logic and statistical rules: base-rate neglect, conjunction fallacy, gambler's fallacy, perception of random sequences, and sample size neglect (Kahneman & Tversky, 1972). Both the anchoring and representativeness scales measured the extent to which participants exhibited biased judgments and decisions relative to normatively, logically, or statistically correct answers. In addition to these bias measures, we also assessed the extent to which participants were able to recognize instances of the three biases before and after the training interventions.

We expected the observational learning intervention to be more effective at debiasing than practice alone (i.e., the control condition). We included the instructional video and serious game interventions as benchmarks by which to assess the efficacy of observational learning. The instructional video served as benchmark for more traditional interventions (e.g., videos and lectures). The serious game intervention provided insight into the potential ceiling for the observational learning intervention, due to the intensive coaching and personalized feedback that it provided to players.

2.1. Methods

2.1.1. Participants and Exclusions.

Three hundred and five people in a convenience sample were recruited through a university participant pool comprised of students, staff, and community residents in Boston, MA (180 women; $M_{\text{age}} = 25.91$, $SD = 10.15$). Each received $30 for completing the experiment. Participants were run in groups of up to 20 at one time in a laboratory on campus. Most participants (82.77%) had at least some college education, 14.52% had a graduate degree, and scored above average on standardized tests of math and verbal ability (median SAT MATH = 650–699; median SAT VERBAL = 600–649). Participants self-identified as White (39.86%), Asian (35.08%), Black (10.81%), or another ethnicity/did not self-identify (14.25%).

Twenty-eight participants failed attention checks ($n = 13$), experienced problems or server and experimenter errors during the study completion ($n = 15$) and were excluded before all the analyses were conducted. As a result, the final sample consists of 277 participants who successfully completed the experiment (Control $n = 74$; Instructional Video $n = 73$; Play Game $n = 66$; Observe Gameplay $n = 64$).

2.1.2. Design

The experiment used a 2 (test battery: pretest, posttest; within-subjects) × 4 (intervention: control, instructional video, play game, observe gameplay; between-subjects) mixed design.

2.1.3. Testing procedure

Each participant sat in a private cubicle with a computer for a 3-hour laboratory session. Participants first completed a battery of pretests, which included scales measuring each of the three biases (i.e., anchoring, social projection, and overreliance on the representativeness heuristic), a five-factor measure of personality (John, Naumann, & Soto, 2008), and the Cognitive Reflection Test (Frederick, 2005). For each of the three bias scales, participants were randomly assigned to complete one of three versions (i.e., version A, B, or C). Immediately after the training interventions (details in the Conditions section below), participants completed a battery of posttests including a different version of each of the three bias scales. Different yet equivalent versions were administered at pretest and posttest (details in the Bias Measures section below).

Participants who received a training intervention (i.e., those in the instructional video, play game, and observe gameplay conditions) answered eleven questions measuring the degree to which they found the training intervention engaging (e.g., the extent to which they found the training intervention interesting, they would like to repeat it in the future, and would recommend it to other people) on 5-point Likert scales (1 = *Strongly Disagree*; 5 = *Strongly Agree*). Participants also provided demographic information.

2.1.4. Conditions

Between the completion of the pretest and posttest measures, participants were randomly assigned (between-subjects) to one of four conditions: control, instructional video, play game, or observe gameplay.

Control. Participants in this condition took a 5-minute break between the batteries of pretests and posttests.

Instructional video. Upon completing the pretest, participants in the Instructional Video condition watched, *Unbiasing Your Biases II,* a 30-minute unclassified training video (Intelligence Advanced Research Projects Activity, 2013).[1] In past research, this game produced medium-sized immediate reductions in each of the three biases (Morewedge et al., 2015). In the video, a narrator first defines heuristics and explains how they can sometimes lead to incorrect inferences. He then defines anchoring, social projection, and the representativeness heuristic, presents vignettes in which actors committed the biases, provides additional examples, and suggests mitigating strategies (e.g., consider alternative anchors, possible outcomes, multiple perspectives, base-rates, and countervailing evidence). A review takes place in the last two minutes of the video.

Play game. Upon completing the pretest, participants in the play game condition played, *Missing—The Final Secret*, a serious game that elicits and mitigates anchoring, social projection, and overreliance on the representativeness heuristic (Symborski et al., 2017). In past research, this game produced large immediate reductions in each of the three biases (Morewedge et al., 2015). In this first person point-of-view educational game, players attempt to exonerate their fictional employer of criminal activity. In three distinct levels, players make judgments designed to test the degree to which they exhibit all three biases during interactive gameplay. Different variants of the three cognitive biases are elicited in each level by requiring players to make in-game decisions based on limited evidence (e.g., estimate the amount of time the average American spends on social media, whether a falling stock is more likely to go up or down the next day, or which of two species of fish is more prevalent in a lake based on small and large catches).

At the end of each level, in an "after-action-review," experts provide definitions and examples of the three biases, players receive personalized feedback on the degree to which they exhibited each bias, and mitigating strategies are suggested. Like the video, the game teaches bias mitigating strategies, including consider alternative explanations, alternative anchors, possible outcomes, different perspectives, base-rates, and countervailing evidence. In addition, the game teaches formal rules of logic (e.g., the conjunction of two events can be no more likely than either event on its own), and basic statistical rules (e.g., small samples yield less reliable population estimates than large samples). Players then are given a chance to practice additional bias eliciting questions. If players give biased answers, they continue to receive additional practice questions (up to 16 in total) and feedback (for more details about the game structure and development, see Barton et al., 2015). Game players took an average of 58 min to complete the game from start to finish.

Observe gameplay. Upon completing the pretest, each participant in this condition was yoked to watch on his or her own computer monitor, one anonymized participant from the play game condition play the debiasing game from beginning to end. The participant's computer mirrored the computer of a game player in real time, including its video and audio output. Computers, monitors, and headsets were identical in the play game and observe gameplay conditions. The mirroring apparatus (devices for monitor and audio signal duplication) were occluded from view. Thus, the participant could not see the game player with

[1] Video available at https://www.youtube.com/watch?v=tPY3xUZvWsY&list=PLfaSGHp0IgDDKoxPpzKw9JvVrozJXxwWt&index=10&t=0s.

whom she was paired or communicate with the game player. The game player was not informed that an observer was watching her gameplay.

2.1.5. Bias measures

Each bias scale measured individual differences in susceptibility to one of three cognitive biases. They were developed by first performing a literature review to identify canonical paradigms used to assess each bias. Multiple variants based on each of those paradigms were then created, which mapped those questions to different subject domains. In order for participants to be tested on different versions of the bias scales at pretest and posttest, we created three interchangeable subscales for each bias (for all versions and items, see Appendices A–C). Scale items in pretests and posttests shared the same problem structures but varied in context and content (see Appendices A–C for lists of all questions). Each participant was randomly administered two of these three subscales, one before and one after the training. The battery of pretests and the battery of posttest scales each took approximately 45 min to complete. Scores on each subscale were averages of item scores that ranged from 0 to 1, with higher scores indicating less biased answers. Overall bias commission scores before and after the training were calculated by averaging the three bias subscale scores (i.e., anchoring, social projection, and representativeness) at that time point. Reliabilities for all the subscales are reported in Table 1.

Anchoring. Fifty-four anchoring questions were developed to measure the tendency to overweight an initially considered piece of information when making a judgment. Forty-two of these items were numerical anchoring items. They asked participants to make numerical estimates in the presence of an anchor that was either provided or self-generated, and either relevant or irrelevant for the judgment at hand (Simmons et al., 2010; Strack & Mussweiler, 1997; Tversky & Kahneman, 1974). Items were scored by calculating the distance between a participant's estimate and the correct answer, divided by the distance between the anchor and the correct answer. These scores varied between 0 and 1, with higher scores indicating less biased answers. Estimates that exceeded the value of the correct answer in the direction away from the anchor were scored as 1. Estimates that exceed the value of the anchor in the direction away from the correct answer were scored as 0 (see Fig. 1).

Twelve additional items assessed focalism by asking participants to forecast their happiness in a hypothetical future situation, and then to revise their initial estimate based on a consideration of the daily activities they would undertake in that situation (Schkade & Kahneman, 1998; Wilson, Lindsey, & Schooler, 2000). Items were scored as the difference between the predicted happiness ratings and the revised happiness ratings. The difference was rescaled to vary between 0 and 1, such that cases in which the revised happiness ratings were equal to the predicted happiness ratings (indicating minimum susceptibility to focalism bias) were scored as 1, and cases in which the two ratings were maximally different (indicating maximum susceptibility to focalism bias) were scored as 0.

The 54 items were split into three subscales of 18 items. Each subscale contained 14 numerical anchoring and 4 focalism items (for all items, see Appendix A).

Social projection. Sixty-nine social projection questions were developed from two paradigms capturing the tendency of judges to overweight their (egocentric) perspective when inferring the beliefs, attitudes, and behaviors of others: the false consensus effect (Ross, Greene, & House, 1977) and attributive similarity (Holmes, 1968; Krueger & Stanke, 2001). The false consensus effect is the tendency to overestimate the extent to which others share similar preferences, thoughts, and behaviors (Ross et al., 1977). Each of 39 false consensus items generally adhered to the format:

Is it a good thing that Supreme Court justices get lifetime appointments? (Yes/No)
What percentage of Americans do you think agrees with your response? (0–100%)

Actual consensus scores were obtained from public opinion polls (e.g., Pew Research Center). For the above item in question, the actual consensus was 33% for yes, 60% for no. Scores varied between 0 and 1 as follows. A participant who either accurately estimated, or underestimated, how many other people agreed with her would receive a score of 1 for this question item as she did not exhibit a false consensus effect. A participant who overestimated how many others agreed with her, however, would exhibit the false consensus effect. In this case, her score would correspond to the difference between her consensus estimate and the actual consensus (as revealed by the poll), divided by the number of possible biased percentage points (i.e., any percentage point above the actual consensus up to 100%). For example, if a participant disapproved of lifetime appointments for Supreme Court justices, and estimated that 92% of Americans agreed with her, she would receive a score of 0.20 [i.e., 1 − (92–60)/(100–60)] for this item. Several questions items asked respondents to indicate what percentage of the population behaves in the same way as them rather than what percentage of the population agrees with them (e.g., What percentage of Americans do you think owns a smartphone?). These were scored in the same way.

Attributive similarity describes a higher likelihood of attributing traits and attitudes that one possesses to others than traits and attitudes that one does not possess (Holmes, 1968; Krueger & Stanke, 2001). Thirty attributive similarity items were created. They adhered to one of two formats:

Attributive similarity format 1:

a. *I don't like to be in situations where people are in disagreement. (1 = strongly disagree; 2 = disagree; 3 = neutral; 4 = agree; 5 = strongly agree)*
b. *The average American doesn't like to be in situations where people are in disagreement. (1 = strongly disagree; 2 = disagree; 3 = neutral; 4 = agree; 5 = strongly agree)*

Attributive similarity format 2:

a. *To what extent do you agree with the following statement? I am adventurous. (1 = not at all; 7 = definitely)*
b. *Imagine you meet another study participant. Do you think that he or she is likely to be adventurous? (1 = not at all; 7 = definitely)*

Answers to each item were scored according to an equation in which the numerator was the absolute value of the difference between the self and other rating, and the denominator was the absolute value of the maximum possible difference between self and other ratings, given the self-rating of the participant (| self-rating – other-rating | / | maximum possible difference score based on self-rating |). For example, if a participant responded "6" to the self-rating question and responded "4" to the other-rating question on the 7-point scale, the numerator would be 2 (= $|6-4|$) and the denominator, the maximum possible difference score, would be 5 (= $|6-1|$, i.e., the absolute value of the difference

Table 1
Cronbach alpha coefficients for bias subscales.

Bias	Subscale	Items	Overall Reliability	Reliability Pretest	Reliability Posttest
Anchoring	A	18	0.62	0.55	0.65
	B	18	0.58	0.48	0.59
	C	18	0.61	0.55	0.47
Projection	A	23	0.83	0.65	0.88
	B	23	0.82	0.59	0.87
	C	23	0.81	0.66	0.84
Representativeness	A	26	0.90	0.82	0.91
	B	26	0.86	0.75	0.89
	C	26	0.88	0.77	0.90

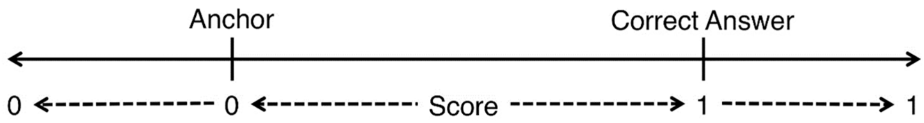

Fig. 1. Scoring of answers by relative distance between the anchor and the correct answer.

between the self-rating and the furthest scale endpoint).

The 69 social projection items were split into three subscales of 23 items each (13 false consensus and 10 attributive similarity items, respectively; for all items, see Appendix B).

Representativeness. Sixty-six representativeness questions were developed from variations of different inappropriate applications of the representativeness heuristic that lead to biased judgments: the conjunction fallacy, the neglect of base rates, the gambler's fallacy, the misperception of randomness, and the neglect of sample sizes (Tversky & Kahneman, 1974). The conjunction fallacy occurs when people believe the likelihood of the co-occurrence of two events is higher than either event on its own [e.g., $p(A \cap B) > p(A)$]. Base-rate neglect occurs when people underweight the general frequency of an event and overweight details of a specific case when predicting its likelihood. Misperceptions of randomness occur when people believe a short run of an event will also represent the characteristics of the full process (e.g., in a coin toss, people think that H-T-H-T-T-H will be more likely than H-H-H-T-T-T, (Tversky & Kahneman, 1974). The gambler's fallacy is a special case of misperceptions of randomness such that after observing a long run of one of two outcomes (e.g., T-T-T-T-T), people believe that the next outcome will reverse the run (e.g., H) to restore the equilibrium. Sample size neglect is the tendency to overestimate the extent to which the frequency of events drawn from small samples would be as predictive as the frequency of events drawn from larger samples.

Sixty items were split into three subscales of 20 items each, and the remaining six items were featured in all three subscales (i.e., each subscale featured 26 items; all items are reported in Appendix C). Statistically or logically correct answers were scored as a 1, and incorrect (biased) answers were scored as a 0.

Bias recognition. Ancillary scales measuring bias recognition were used to assess the ability of participants to recognize and discriminate between instances of the three biases. Fifty-seven items were developed and were split into three subscales of 19 items each. These items described an instance in which a person exhibited one of the three biases, and participants indicated which bias they believed to have been exhibited in that instance in a multiple-choice format. Participants received a score of one for choosing the correct answer and zero for choosing an incorrect answer. Bias recognition scales were averaged and then rescaled to a 0–100 scale, with higher scores indicating greater ability to recognize and discriminate between the three biases.

2.2. Results

All statistical tests reported in this paper are two-tailed, unless noted otherwise. Prior to the main analyses, bias scores were rescaled to a 0–100 scale where 0 denotes completely biased judgments and 100 denotes completely unbiased judgments. Our analysis code and data are posted at https://osf.io/9cujz/. In Experiment 1, the significance levels (*p*-values) of comparisons from pretest to postest are uncorrected; the significance levels (*p*-values) of comparisons across conditions at posttest have undergone Bonferroni correction.

2.2.1. Scale reliability

The internal consistency (Cronbach's α) of the bias subscales was comparable to or higher than those observed for other decision making scales (e.g., A-DMC, Bruine De Bruin et al., 2007) for Anchoring ($M_{\alpha\ pretest} = 0.54$; $M_{\alpha\ posttest} = 0.57$), Social Projection ($M_{\alpha\ pretest} = 0.63$; $M_{\alpha\ posttest} = 0.86$) and Representativeness ($M_{\alpha\ pretest} = 0.85$; $M_{\alpha\ posttest} = 0.93$). Reliabilities for all the individual bias subscales are reported in Table 1.

2.2.2. Overall bias reduction

Averaging across all three biases, a 2 (test battery: pretest vs. posttest) × 4 (intervention: control, instructional video, play game, and observe gameplay) mixed ANOVA revealed a main effect of test battery, $F(1, 273) = 643.77$, $p < .001$, $\eta_p^2 = 0.70$, a main effect of intervention, $F(3, 273) = 20.47$, $p < .001$, $\eta_p^2 = 0.18$, and a test battery × intervention interaction, $F(3, 273) = 56.36$, $p < .001$, $\eta_p^2 = 0.38$. Paired sample *t-tests* comparing pretest to posttest scores within conditions using Bonferroni correction revealed that in all four conditions participants exhibited significant reductions in overall cognitive bias. Controls, who had only gained the benefit of practice from taking the pretest, exhibited a small bias reduction at posttest, $t_{Control} = 3.24$, $p = .007$, $d_z = 0.38$. By contrast, training based interventions (the observe gameplay, instructional video, and play game conditions) yielded large overall reductions in cognitive bias at posttest relative to pretest, $t_{ObserveGameplay} = 14.89$, $p < .001$, $d_z = 1.86$; $t_{InstructionalVideo} = 13.01$, $p < .001$, $d_z = 1.52$; $t_{PlayGame} = 16.29$, $p < .001$, $d_z = 2.00$ (for all means, see Table 2).

We compared debiasing across participants in all four conditions at posttest with ANCOVA, using pretest bias scores as a covariate. This analysis yielded a significant main effect of intervention, $F(3, 272) = 63.48$, $p < .001$, $\eta_p^2 = 0.41$. Tukey post hoc tests using the covariate adjusted mean showed that the training based interventions were more effective than the control condition, $ts > 7.73$, $ps < 0.001$, $ds > 1.27$. The observe gameplay condition was more effective than the instructional video condition, $t = 2.87$, $p = .022$, $d = 0.49$, but not significantly different from the play game condition, $t = 2.47$, $p = .066$. The play game condition was also more effective than the instructional video condition, $t = 5.43$, $p < .001$, $d = 0.92$. We next examined the debiasing effects of the interventions with respect to each of the three specific biases they ameliorated.

Anchoring. A 2 (test battery: pretest vs. posttest) × 4 (intervention: control, instructional video, play game, and observe gameplay) mixed ANOVA showed a significant main effect of test battery, $F(1, 273) = 135.01$, $p < .001$, $\eta_p^2 = 0.33$, a main effect of intervention, $F(3, 273) = 6.04$, $p < .001$, $\eta_p^2 = 0.06$, and a test battery × intervention interaction, $F(3, 273) = 12.90$, $p < .001$, $\eta_p^2 = 0.12$. Paired sample *t-tests* comparing pretest to posttest scores using Bonferroni correction demonstrated that the training interventions produced large reductions in anchoring, $t_{ObserveGameplay} = 6.56$, $p < .001$, $d_z = 0.82$; $t_{InstructionalVideo} = 6.96$, $p < .001$, $d_z = 0.81$; $t_{PlayGame} = 8.61$, $p < .001$, $d_z = 1.06$. Practice alone (i.e., the control condition) did not reduce anchoring, $t < 1$ (Table 2).

We compared debiasing across participants in all four conditions at posttest with ANCOVA, including pretest scores as a covariate. This analysis yielded a significant main effect of intervention, $F(3, 272) = 18.56$, $p < .001$, $\eta_p^2 = 0.17$. Tukey post hoc tests using the covariate adjusted mean showed that the training based interventions were more effective than the control condition, $ts > 4.31$, $ps < 0.001$, $ds > 0.71$. The observe gameplay condition, however, was not significantly more effective than the instructional video condition, $t = 0.75$, $p = .876$, or the play game condition, $t = 2.20$, $p = .126$. The play game condition was more effective than the instructional video condition, $t = 3.03$, $p = .014$, $d = 0.51$.

Social projection. A 2 (test battery: pretest vs. posttest) × 4 (intervention: control, instructional video, play game, and observe gameplay) mixed ANOVA showed a significant main effect of test battery, $F(1, 273)$

Table 2
Debiasing effects of training interventions on bias commission and recognition.

	Control (Practice)	Instructional Video (Practice + Information)	Play Game (Practice + Information + Feedback)	Observe Gameplay (Practice + Observation)
Overall Bias Commission				
Pretest	45.79_a (9.40)	46.20_a (8.40)	48.75_a (8.77)	46.99_a (8.62)
Posttest	48.09_a (10.45)*	58.25_b (10.81)*	67.38_c (10.68)*	62.66_c (8.95)*
Anchoring Bias				
Pretest	36.49_a (12.33)	36.25_a (9.82)	37.06_a (10.47)	35.94_a (10.33)
Posttest	37.78_a (12.61)	44.78_b (11.42)*	50.25_c (9.68)*	45.94_{bc} (9.27)*
Social Projection				
Pretest	47.43_a (8.47)	48.37_a (8.29)	49.59_a (8.69)	48.85_a (9.09)
Posttest	48.48_a (9.32)	52.41_b (9.07)*	66.68_c (13.96)*	60.97_d (10.55)*
Representativeness				
Pretest	53.45_a (19.61)	53.97_a (16.87)	59.60_a (18.02)	56.17_a (17.27)
Posttest	58.00_a (21.76)*	77.57_b (19.65)*	85.20_b (17.97)*	81.06_b (17.09)*
Bias Recognition				
Pretest	28.20_a (17.86)	31.19_a (15.65)	33.38_a (20.60)	32.40_a (16.27)
Posttest	30.73_a (19.40)	72.25_b (21.79)*	61.99_c (24.99)*	57.78_c (23.50)*

Note: Scales range = 0–100, with lower values indicating more biased answers. Standard deviations are in parentheses. Means within rows that do not share a common subscript differ at $p < .05$ (Tukey's post hoc comparisons). Asterisks indicate a significant bias reduction between Pretest and Posttest in each column, all $ps < 0.01$*. Posttest comparisons include pretest scores as a covariate.

= 169.26, $p < .001$, $\eta_p^2 = 0.38$, a main effect of intervention, $F(3, 273) = 21.90$, $p < .001$, $\eta_p^2 = 0.19$, and a test battery × intervention interaction, $F(3, 273) = 31.28$, $p < .001$, $\eta_p^2 = 0.26$. Paired sample t-tests comparing pretest to posttest scores using Bonferroni correction revealed that the training interventions providing information, personalized feedback, or observation produced medium to large reductions in social projection, $t_{\text{ObserveGameplay}} = 7.95$, $p < .001$, $d_z = 0.99$; $t_{\text{InstructionalVideo}} = 3.84$, $p = .001$, $d_z = 0.45$; $t_{\text{PlayGame}} = 9.80$, $p < .001$, $d_z = 1.21$. Practice alone (i.e., the control condition) did not reduce social projection, $t = 1.15$, $p = 1.00$ (Table 2).

We compared debiasing across participants in all four conditions at posttest with ANCOVA, including pretest scores as a covariate. This analysis yielded a significant main effect of intervention, $F(3, 272) = 41.86$, $p < .001$, $\eta_p^2 = 0.32$. Tukey post hoc tests using the covariate adjusted mean showed that both the play game and the observe gameplay conditions were more effective than the control condition, $ts > 6.90$, $ps < 0.001$, $ds > 1.18$. The instructional video condition was not significantly more effective than the control condition, $t = 2.10$, $p = .155$. The observe gameplay condition was more effective than the instructional video condition, $t = 4.86$, $p < .001$, $d = 0.83$, but less effective than the play game condition, $t = 3.05$, $p = .014$, $d = 0.53$. The play game condition was more effective than the instructional video condition, $t = 8.04$, $p < .001$, $d = 1.37$.

Representativeness. A 2 (test battery: pretest vs. posttest) × 4 (intervention: control, instructional video, play game, and observe gameplay) mixed ANOVA revealed a significant main effect of test battery, $F(1, 273) = 464.84$, $p < .001$, $\eta_p^2 = 0.63$, a main effect of intervention, $F(3,273) = 12.41$, $p < .001$, $\eta_p^2 = 0.12$, and a test battery × intervention interaction, $F(3,273) = 32.05$, $p < .001$, $\eta_p^2 = 0.26$. Paired sample *t-tests* comparing pretest to posttest scores using Bonferroni correction revealed that in all four conditions, participants reduced their overreliance on the representativeness heuristic, $t_{ObserveGameplay} = 12.32$, $p < .001$, $d_z = 1.54$; $t_{Control} = 3.22$, $p = .008$, $d_z = 0.37$; $t_{InstructionalVideo} = 12.07$, $p < .001$, $d_z = 1.41$; $t_{PlayGame} = 13.57$, $p < .001$, $d_z = 1.67$ (Table 2).

We compared debiasing across participants in all four conditions at posttest with ANCOVA, including pretest scores as a covariate. This analysis yielded a significant main effect of intervention, $F(3, 272) = 39.12$, $p < .001$, $\eta_p^2 = 0.30$. Tukey post hoc tests using the covariate adjusted mean showed that all three training based interventions (play game, observe gameplay, and instructional video conditions) were more effective than the control condition, $ts > 8.12$, $ps < 0.001$, $ds > 1.34$. There were no significant differences in bias reduction among the three training based interventions, $ts < 1.46$, $ps > 0.46$.

2.2.3. Bias recognition

A 2 (test battery: pretest vs. posttest) × 4 (intervention: control, instructional video, play game, and observe gameplay) mixed ANOVA revealed a significant main effect of test battery, $F(1, 273) = 386.02$, $p < .001$, $\eta_p^2 = 0.59$, a main effect of intervention, $F(3, 273) = 23.09$, $p < .001$, $\eta_p^2 = 0.20$, and a test battery × intervention interaction, $F(3, 273) = 44.57$, $p < .001$, $\eta_p^2 = 0.33$. Paired sample *t-tests* comparing pretest to posttest scores using Bonferroni correction revealed only training based interventions produced improvements in bias recognition, $t_{ObserveGameplay} = 10.04$, $p < .001$, $d_z = 1.26$; $t_{InstructionalVideo} = 14.98$, $p < .001$, $d_z = 1.75$; $t_{PlayGame} = 10.93$, $p < .001$, $d_z = 1.35$. Participants in the control condition did not exhibit improvements in bias recognition, $t_{Control} = 1.27$, $p = .835$, $d_z = 0.15$ (Table 2).

We compared bias recognition across all four conditions at posttest with ANCOVA, including pretest scores as a covariate. This analysis yielded a significant main effect of intervention, $F(3, 272) = 53.01$, $p < .001$, $\eta_p^2 = 0.37$. Tukey post hoc tests using the covariate adjusted mean showed that the training based interventions were more effective than the control condition, $ts > 7.29$, $ps < 0.001$, $ds > 1.25$. Among the three training based interventions, the instructional video condition better taught bias recognition than the play game, $t = 3.50$, $p = .003$, $d = 0.59$, and the observe gameplay conditions, $t = 4.55$, $p < .001$, $d = 0.78$. The observe gameplay condition was not significantly different from the play game condition, $t = 1.05$, $p = .720$.

Providing some insight into the mechanism by which debiasing occurred, participants who became better at bias recognition from the pretest to the posttest demonstrated more bias reduction from the pretest to the posttest. Indeed, there were positive correlations between improved bias recognition and bias reduction from pretest to posttest among the participants who received any of the three training based interventions, $r_{PlayGame} = 0.28$, $p = .026$, $r_{ObserveGameplay} = 0.35$, $p = .004$, and $r_{InstructionalVideo} = 0.42$, $p < .001$. Participants in the control condition did not exhibit such a relationship, $r_{Control} = -0.14$, $p = .224$.

2.2.4. Engagement

A one-factor ANOVA revealed that participants reported different levels of engagement with the instructional video, play game, and observe gameplay interventions, $F(2, 188) = 11.96$, $p < .001$, $\eta_p^2 = 0.11$. Tukey post hoc analysis revealed that engagement with the observe gameplay and instructional video interventions did not differ ($M_{ObserveGameplay} = 3.18$, $SD_{ObserveGameplay} = 0.64$; $M_{InstructionalVideo} = 2.99$, $SD_{InstructionalVideo} = 0.37$; $p = .100$), but both interventions were less engaging than the play game intervention ($M_{PlayGame} = 3.43$, $SD_{PlayGame} = 0.56$, all $ps < 0.030$). Across the three interventions, controlling for overall pretest scores, engagement with the training intervention was weakly correlated with overall posttest scores, $r_{partial} = 0.18$, $p = .012$.

2.2.5. Additional analyses

We report the results of additional exploratory analyses that may be of interest for the development of future debiasing interventions using observational learning. We first examined whether participants in the observe gameplay condition exhibited greater debiasing if they watched a game player who exhibited more or less bias during gameplay. There was no relationship, $r = 0.04$, $p = .775$. In addition, there was no relationship between bias reduction from pretest to posttest within a game player and the participant observing that player, $r = 0.09$, $p = .533$. These results suggest that the effects of the observational learning-based intervention were not contingent on player performance or improvement within the player observed.

We also examined whether any factors predicted which participants would benefit the most from the observational learning-based training intervention. To this end, we examined the correlations between demographic and engagement variables with overall bias scores at pretest (Table 3, left column) and at posttest (i.e., partial correlations, while controlling for overall bias at pretest; Table 3, right column). Age was the only significant predictor after appropriate Bonferroni corrections. Older participants benefitted more from observational learning than did younger participants, perhaps because they tended to be more biased (at pretest) and had more potential to improve (at posttest).

2.3. Discussion

Overall, an observational learning-based intervention had a large immediate debiasing effect ($d_z = 1.86$) on the extent to which participants exhibited three cognitive biases: anchoring, social projection, and representativeness (all individual $ds \geq 0.82$). Observational learning was effective in the sense that it reduced commission of all three cognitive biases compared to a control condition in which participants received no intervention other than practice from having taken the pretests. Overall, the observational learning-based intervention was also significantly more effective than the instructional video intervention. It is worth noting that the comparison between the efficacy of the instructional video and play game conditions also directly replicates previous research (Morewedge et al., 2015), where the two interventions had large immediate debiasing effects (all $ds > 1.52$), but the game had a larger debiasing effect than did the video. We speculate that the positive relationship between improvements in bias recognition and bias reduction suggests that the interventions improved judgment by teaching how to identify when bias might influence judgment, and which bias reducing strategies to apply in those cases.

Table 3
Demographic correlates of bias commission and mitigation in observe gameplay condition.

	Initial Bias Commission (Pretest)	Bias Mitigation (Posttest, Controlling for Pretest)
Overall Bias Commission (Posttest)	0.448***	–
Engagement	−0.133	−0.148
Cognitive Reflection Test	0.198	−0.234
SAT Verbal	0.394**	0.021
SAT Math	0.282*	−0.207
Education Level	0.067	0.155
GPA	−0.097	0.08
Household Income	0.075	−0.08
Age	−0.342**	0.419***
Gender	0.092	0.019
Native English Speaker	−0.077	−0.224
Big 5 Personality Traits		
###Agreeableness	−0.179	−0.068
###Conscientiousness	−0.003	0.259*
###Extraversion	−0.088	0.02
###Neuroticism	0.021	−0.283*
###Openness	0.016	−0.101

Note: Initial bias indicates correlation (r) between individual differences and overall bias commission at pretest for participants in conditions receiving a training intervention (instructional video, play game, observe gameplay). Bias Mitigation indicates partial correlation ($r_{partial}$) between individual differences and overall bias commission at posttest controlling for pretest scores. Asterisks indicate statistical significance, $p < .05^*$, $p < .01^{**}$, $p < .001^{***}$.

Of course, these conclusions are limited by the fact that participants in the observational learning condition also received the same information as did participants in the play game condition, which was conceptually similar in its content to the instructional video but different in its execution. In other words, the examples and bias definitions used in the observe gameplay and play game conditions were different than the examples and bias definitions in the instructional video condition. It is possible that the different presentation of that information drove the differences between the instructional video and observe gameplay conditions.

In Experiments 2A and 2B, we directly addressed this potential confound. We examined the effectiveness of observational learning on its own by completely isolating, in two separate conditions, an information-based intervention and an observational learning intervention. Participants were either taught a decision rule or saw another participant using the rule. We also included a condition in which participants received the combination of both interventions. This design allowed us to test the independent contributions of both debiasing interventions (i.e., observational learning and instruction) and whether they would have additive effects. The training interventions tested in Experiments 2A and 2B were also much easier to scale than the interventions tested in Experiment 1. Each took about five minutes to administer.

3. Experiments 2A and 2B

In the context of a Weight on Advice (WOA) paradigm (Gino & Moore, 2007; Gino, 2008; Harvey & Fischer, 1997; Yaniv, 2004), we examined the ability of a brief observational learning intervention to teach participants a simple decision rule—the averaging principle. Judgments are generally improved when judges average their own inferences with the advice of another person (Larrick & Soll, 2006; Soll & Larrick, 2009). We examined the efficacy of an observational learning intervention that taught this averaging rule, placed between two sets of weight estimates. Most important, we compared it to the effects of practice alone (i.e., our control condition). In addition, in a factorial design, we tested how it worked by comparison, and in conjunction with, an orthogonal information-based intervention. This allowed us to examine the combined effects of observational learning and information-based interventions. Given the benefits of having both explicit knowledge of abstract rules and a model of their implementation in other domains (e.g., Kappes & Morewedge, 2016; Zimmerman & Rosenthal, 1974), we expected their effects on decision rule learning to be additive. We expected participants who received the combined intervention to increase their use of the averaging rule more than participants who received the information-based intervention.

The WOA paradigm assesses the extent to which judges incorporate advice by comparing their initial estimates, advice received, and revised estimates. Using modified and classic formulas to extract WOA, we assessed the extent to which the training interventions effectively taught the averaging rule and increased advice taking. In each of the two experiments, we also explored the potential range of benefits to accuracy provided by the training interventions when advice was either naturalistic or optimized to maximize accuracy if participants followed the averaging rule. In Experiment 2A we examined the effects of the interventions in a naturalistic scenario by providing participants with representatively sampled advice. In Experiment 2B we examined their effects when the advice was provided by an algorithm, portrayed as another participant, that generated unique optimized advice for each initial estimate. The algorithm made sure that, if participants averaged their initial estimate with this optimized advice, their final estimate would always be closer to the true value. Other minor procedural differences between the two experiments are discussed in the methods section.

In our analyses of both experiments, we examined the extent to which participants learned the averaging rule (i.e., our measure of learning) and their reliance on advice increased from pretest to posttest. In both experiments, we expected observational learning to outperform the control condition on learning and weight on advice, regardless of the quality of advice provided. Furthermore, we expected the effects of observational learning and information-based interventions on rule learning and advice taking to be additive.

We also examined the extent to which the interventions improved the objective accuracy of participants' estimates from pretest to posttest, both in percentage and in absolute terms. We were more agnostic about the ability of the interventions to improve accuracy, and expected advice quality to be more important for these analyses than for learning and advice taking. We expected any accuracy improvements conferred by the observational learning intervention, relative to the control condition, to be larger when advice was optimized than when advice was naturalistic (Experiments 2B and 2A, respectively).

3.1. Methods

3.1.1. Participants

Experiment 2A. We recruited 1000 participants on Amazon Mechanical Turk, 998 accessed the survey; 980 completed the experiment without technical errors such as experiencing technical issues or inability to watch the videos in the two conditions involving observational learning; if they provided incomplete responses; or if they provided duplicate responses from the same participant ID. They received $2 as compensation. We excluded 84 participants who made three or more judgments (out of the 40 total) in excess of three times the interquartile range of each estimate (Tukey's fence rule; 3*IQR), resulting in a final sample size of 896 participants (59.70% male, 39.50% female, 6 participants preferred not to report their gender; $M_{\mathrm{age}} = 37.89$, $SD_{\mathrm{age}} = 11.23$). All participant exclusions were determined before the analyses were performed; there were no other exclusions. Analysis code and data are available at https://osf.io/9cujz/.

Experiment 2B. We recruited 505 participants on Amazon Mechanical Turk; 494 completed the experiment without technical errors, duplicate responses, or incomplete responses (55.47% male, 44.33% female, one participant preferred not to report his/her gender; $\mathrm{M}_{\mathrm{age}} = 37.77$, $\mathrm{SD}_{\mathrm{age}} = 10.77$) and received $2 as compensation. All participant exclusions were determined before analyses were performed; there were no other exclusions. Analysis code and data are available at https://osf.io/9cujz/. The application of Tukey's fence rule to Experiment 2A was due to learnings distilled from running Experiment 2B, which was run first.

3.1.2. Design

The experiments used a 2 (WOA: set 1, set 2; within-subjects) × 2 (observational learning intervention, yes, no; between-subjects) × 2 (information-based intervention: yes, no; between-subjects) mixed design.

3.1.3. Estimates

Participants estimated the weight of two sets of 10 household objects (e.g., a teapot, a cutting board, a bar stool), 20 household objects in total. Objects were presented one at a time. Participants saw an image of the object and its dimensions (i.e., length, width, and height). After making an initial weight estimate, participants were shown another weight estimate (i.e., the advice), described as the estimate made by another participant for that object. They were next prompted to submit a revised weight estimate for that object. After submitting this revised estimate, participants were immediately told the true weight of the object. Objects in Set 1 were presented in the same (randomly) predetermined order for all participants. Objects in Set 2 were presented in a random order (i.e., order varied across participants).

3.1.4. Advice

Naturalistic. In Experiment 2A we aimed to create a naturalistic scenario in terms of advice provided, such that the advice offered would be representative of weights estimates for the 20 objects. To this end, the advice provided was the median of the initial weight estimates provided by participants in Experiment 2B (which was run first). This approach did not guarantee that implementation of the averaging rule would always improve the accuracy of each estimate in Set 2.

Optimized. In Experiment 2B we used an algorithm to provide optimized advice. This optimized advice always bracketed the correct weight of the object, and its value was dynamically calculated during the experiment, using the participant's initial weight estimate and the true weight of the object. Following the averaging rule would, thus, always improve the accuracy of the revised estimate. Advice was calculated using the following formula:

$$\begin{aligned} Optimized\ \ Advice = {} & 2\cdot(\text{true weight} + \text{jittering}) \\ & - \text{participant's initial weight estimate} \end{aligned}$$

The jittering was performed by randomly adding noise to the dynamic advice. That noise ranged from ±0.05 lbs. to ±0.15 lbs. If a participant relied on the averaging rule using the dynamic advice above, her revised estimate should have been within ±0.15 lbs. of the true weight of the object.

3.1.5. Training interventions

Each participant was randomly assigned to a training intervention comprised of one of four combinations of the observational learning (No vs. Yes) and information-based interventions (No vs. Yes): Control, Information (No, Yes), Observational Learning (Yes, No), or Information and Observational Learning (Yes, Yes).

Control. After providing the weight estimates of objects in Set 1, participants in this condition directly proceeded to estimate the weight of the objects in Set 2.

Information. After providing the weight estimates of objects in Set 1, participants in this training condition read an explanation of the benefits of averaging estimates in improving accuracy, adapted from Larrick and Soll (2006). The explanation differed between Experiment 2A and Experiment 2B to ensure consistency between the information and the nature of the advice provided.

Experiment 2A

*"Before you continue with the next set of 10 estimates, we will provide you with some information on how you can effectively use the estimate of the other participant. Averaging estimates is an effective way to improve the overall accuracy of sets of quantitative judgments, like the estimates you are making in this study. If your estimate and another participant's estimate fall on different sides of the correct value (*i.e., *they 'bracket' the true value), then the average of the two estimates is usually closer to the correct value than your original estimate. If your estimate and their estimate fall on the same side of the correct value, then the average of the two estimates will often be closer to the correct value than your original*

estimate. In other words, even though both of your and their estimates might be imperfect, averaging estimates improves accuracy in the majority of cases. When receiving someone else's advice about quantity estimates, then, an effective way to improve the overall accuracy of your set of estimates is to take the average of their "advice estimate" and your original estimate."

Experiment 2B

*"Before you continue with the next set of 10 estimates, we will provide you with some information on how to effectively use the estimate of the other participant. Averaging estimates is an effective way to improve the accuracy of quantitative judgments such as the estimates you are making in this study. If your estimate and another participant's estimate fall on different sides of the correct value (*i.e.*, they 'bracket' the true value), then their average is always going to be closer to the correct value than each of your and the other participant's original estimates. When receiving someone else's advice about a quantity estimate, an effective way to improve the accuracy of your estimate is to take the average of their "advice estimate" and your original estimate."*

Observational learning. After providing the weight estimates of objects in Set 1, participants in the observational learning condition watched a screencast (i.e., screen capture video) of another "participant" (who was actually a research assistant) making weight estimates for a subset of objects from Set 1. For each object, the screencast first showed the object and the initial weight estimate as the "actor" typed it into the response box. Next, participants saw the advice the actor received and saw her averaging it with her own initial estimate using a computer calculator (i.e., the two estimates were added together and then divided by two). The actor then entered that average in the response box, submitting it as her revised estimate.

In Experiment 2A ("naturalistic advice"), participants saw the actor complete the five estimation tasks (with a total duration of approximately 5 min) from the first set of object weight estimates, which varied on whether the original estimate and the advice bracketed the true value or not, and whether averaging would improve the accuracy of the estimate. Participants did not see any accuracy feedback provided to the actor in the screencasts shown in Experiment 2A. They could not see whether initial estimates were improved by use of the averaging rule.

In Experiment 2B ("optimized advice"), participants saw the actor complete the same five estimation tasks (with a total duration of approximately 5 min), but in all cases the original estimate and the advice bracketed the true value, and averaging improved the final estimate. In Experiment 2B, participants also saw a comparison of the actor's initial estimates, advice, revised estimates, and the objects' true weight. Thus, they saw that her initial estimates were always improved by use of the averaging rule.

Information plus observational learning. After providing the weight estimates of objects in Set 1, participants in this training condition were first administered the information-based training intervention and then the observational learning-based intervention before they proceeded to estimate the weight of the objects in Set 2.

3.1.6. Dependent measures

Advice taking. Our first dependent measure captured the extent to which participants learned and implemented the averaging rule. To calculate it, we relied on a modified version of the typical WOA formula (Gino & Moore, 2007; Gino, 2008; Harvey & Fischer, 1997; Yaniv, 2004). The WOA formula assesses the extent to which participants incorporate the advice received into their revised weight estimates:

$$WOA = \frac{revised\ \ estimate - initial\ \ estimate}{advice - initial\ \ estimate}$$

In this formula, weight on advice was equal to 0 when participants made no revision to their estimate after receiving the advice—when their revised estimate was equal to their initial estimate. Weight on advice was equal to 1 when participants replaced their revised estimate with the advice; this is the maximum relative weight that can be given to the advice received. Weight on advice was equal to a value between 0 and 1 when the revised estimate combined the initial estimate and advice received. In cases where the revised estimate did not fall between the initial estimate and the advice received (e.g., participants updated their revised estimates in a direction away from the advice they received), so the weight on advice was a negative number, it was adjusted to 0.

Our primary goal was to measure to what extent participants implement the averaging rule. Thus, our modified WOA formula measured how close participants' revised estimate was to the average between the initial estimate and the advice. This measure was equal to 1 if the revised estimate was exactly equal to the average between the initial estimate and the advice, indicating maximum learning; it was equal to 0 in case the revised estimate was equal to either the initial estimate or the advice indicating minimum learning, and would otherwise assume a value equal to:

$$Learning = -|2{\cdot}WOA - 1| + 1$$

to capture the difference between the revised estimate and the average between the initial estimate and the advice.

Accuracy. We also measured the effectiveness of the interventions at improving accuracy by comparing each weight estimate to the correct weight of the object. The estimation bias, the difference between the estimate and the true weight of each object, was calculated as:

$$Estimation\ \ Bias = \ln\left(\frac{|weight\ \ estimate - true\ \ weight|}{true\ \ weight} + 1\right)$$

Estimation bias was equal to 0 when a weight estimate was equal to the true weight of the object. It increased as the participant's estimate deviated from the true weight. We applied a natural log-transformation to account for skewness in estimates. Based on this estimation bias formula, we created two dependent measures. First, we calculated the *percentage of improved judgments*, the number of estimates that became more accurate upon the incorporation of the advice (i.e., those estimates for which estimation bias$_{initial\ \ estimate}$ > estimation bias$_{revised\ \ estimate}$). Second, we calculated the *absolute improvement* of the estimates, whether participants became more accurate after incorporating advice (i.e., estimation bias$_{initial\ estimate}$ - estimation bias$_{revised\ estimate}$). A positive value indicates that estimation bias decreased from the initial to the revised estimates; a negative value indicates that estimation bias increased from the initial to the revised estimates.

3.2. Results

3.2.1. Rule learning

For both experiments, we first compared the extent to which interventions increased the use of the averaging rule from pretest to posttest. For all means, see Fig. 2.

Experiment 2A. As our primary measure of learning, we compared use of the averaging rule across the four training intervention conditions at posttest with a 2 (Observation: Yes vs. No) × 2 (Information: Yes vs. No) ANCOVA, using pretest scores as a covariate. This analysis yielded a significant main effect of observational learning, $F(1, 891) = 67.67$, $p < .001$, $\eta_p^2 = 0.07$, and a significant main effect of information, $F(1, 891) =$

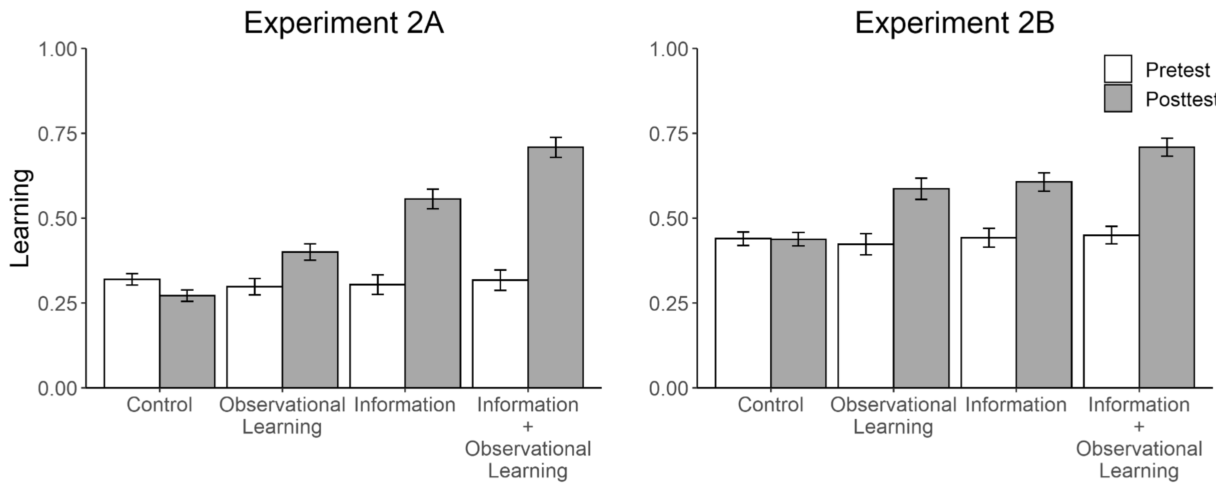

Fig. 2. Use of the averaging rule at pretest and posttest by debiasing training intervention, in Experiments 2A (left panel) and 2B (right panel). Scale ranges from 0 to 1; error bars represent 95% CIs.

292.06, $p < .001$, $\eta^2_p = 0.25$. The interaction effect was not significant, $F(1, 891) = 0.02$, $p = .875$. A simple effect analysis revealed that participants in the observational learning condition used the averaging rule more at posttest than did participants in the control condition, $t(891) = 5.72$, $p < .001$, $d = 0.54$. Indicating additivity in training effects on learning, participants who received the combined interventions (i.e., the information plus observational learning condition) used the averaging rule more at posttest than did participants who only received information-based training, $t(891) = 5.91$, $p < .001$, $d = 0.56$.

Experiment 2B. We conducted parallel analyses for Experiment 2B. We compared use of the averaging rule across the four training intervention conditions at posttest with a 2 (Observation: Yes vs. No) × 2 (Information: Yes vs. No) ANCOVA, using pretest scores as a covariate. This analysis yielded a significant main effect of observational learning, $F(1, 489) = 55.07$, $p < .001$, $\eta^2_p = 0.10$, and a significant main effect of information, $F(1, 489) = 65.04$, $p < .001$, $\eta^2_p = 0.12$. The interaction was not significant, $F(1, 489) = 2.72$, $p = .100$. Replicating the results of Experiment 2A, a simple effect analysis revealed that participants in the observational learning condition used the averaging rule at posttest more than did participants in the control condition, $t(489) = 6.46$, $p < .001$, $d = 0.82$. Indicating additivity in training effects on learning, participants who received the combined interventions used the averaging rule more at posttest than did participants who only received the information-based intervention, $t(489) = 4.05$, $p < .001$, $d = 0.52$.

Together, these results provide evidence that observational learning based debiasing interventions can teach a simple decision rule, and that their effect is additive with respect to information-based debiasing interventions.

3.2.2. *Weight on advice*

We next examined how the interventions influenced the absolute weight participants put on the advice they received. For all means, see Fig. 3.

Experiment 2A. We compared WOA across the four training intervention conditions at posttest with a 2 (Observation: Yes vs. No) × 2 (Information: Yes vs. No) ANCOVA, using pretest scores as a covariate. This analysis yielded a significant main effect of observational learning, $F(1, 891) = 27.46$, $p < .001$, $\eta^2_p = 0.03$, a significant main effect of information, $F(1, 891) = 143.95$, $p < .001$, $\eta^2_p = 0.14$, and a significant interaction, $F(1, 891) = 4.27$, $p = .039$, $\eta^2_p = 0.005$. A simple effect

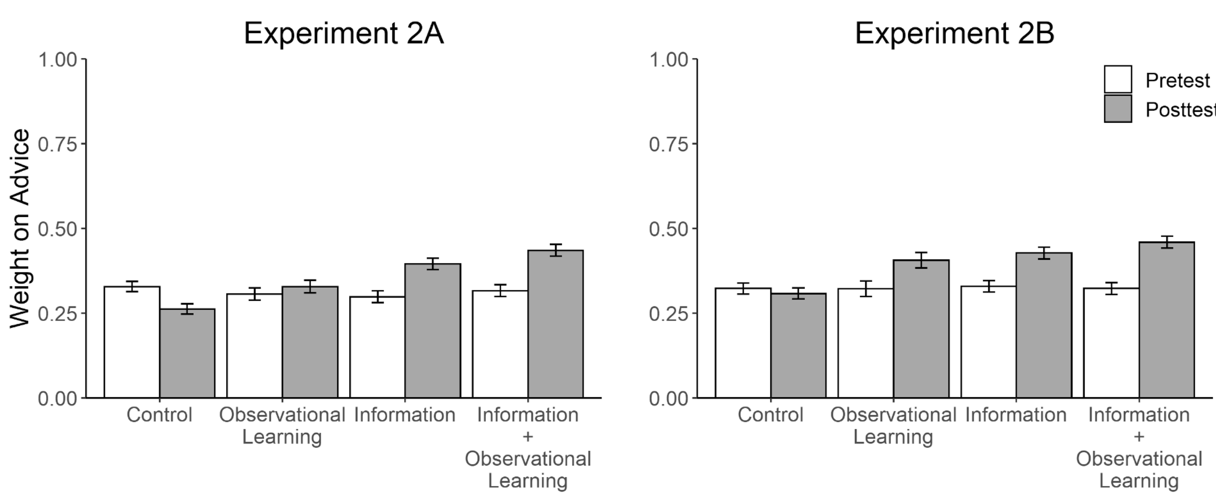

Fig. 3. Weight on advice at pretest and posttest by debiasing training intervention, in Experiments 2A (left panel) and 2B (right panel). Scale ranges from 0 to 1; error bars represent 95% CIs.

analysis revealed that participants in the observational learning condition gave more weight to advice than did participants in control condition, $t(891) = 5.18$, $p < .001$, $d = 0.49$. Indicating additivity in training effects on learning, participants who received the combined interventions gave more weight to advice than did participants who only received the information-based training intervention, $t(891) = 2.23$, $p = .026$, $d = 0.21$.

Experiment 2B. We conducted parallel analyses for Experiment 2B. We compared WOA across the four training intervention conditions at posttest with a 2 (Observation: Yes vs. No) × 2 (Information: Yes vs. No) ANCOVA, using pretest scores as a covariate. This analysis yielded a significant main effect of observational learning intervention, $F(1, 489) = 34.51$, $p < .001$, $\eta_p^2 = 0.07$, and a significant main effect of information intervention, $F(1, 489) = 55.20$, $p < .001$, $\eta_p^2 = 0.10$, and a significant interaction, $F(1, 489) = 7.83$, $p = .005$, $\eta_p^2 = 0.02$. Replicating the results of Experiment 2A, a simple effect analysis revealed that participants in the observational learning condition gave more weight to advice than did participants in the control condition, $t(489) = 6.18$, $p < .001$, $d = 0.78$. Indicating additivity in training effects on learning, participants who received the combined interventions gave more weight to advice than did participants who only received the information-based training intervention, $t(489) = 2.16$, $p = .031$, $d = 0.28$.

Together, these results provide evidence that observational learning based debiasing interventions lead people to give more weight to advice, and that their effect is additive with respect to information-based debiasing interventions.

3.2.3. Accuracy gains

We then examined the extent to which the interventions improved the accuracy of estimates from pretest to posttest. We analyzed first the percentage of judgments for which the revised weight estimate became closer to the actual weight of the object, and then compared the overall reduction in estimation bias (for all means, see Fig. 4). Note that the nature of the advice in Experiments 2A and 2B was different. In Experiment 2A, implementing the averaging rule did not guarantee greater accuracy. Only 25.12% of the advice in Experiment 2A bracketed the true weight, which imposed somewhat of a ceiling on the potential for the rule to improve accuracy. If participants fully implemented the averaging rule for all revised estimates in set 2, we would have anticipated improvements in only 66.38% of all revised estimates. By contrast, in Experiment 2B, implementing the averaging rule guaranteed greater accuracy. The dynamic advice *always* bracketed the true weight, so averaging would always have led to greater accuracy. This difference was pronounced in its effects across both the percentage of improved weight judgments and their absolute accuracy.

Experiment 2A. We first compared the percentage of revised estimates that improved at posttest across the four conditions with a 2 (Observation: Yes vs. No) × 2 (Information: Yes vs. No) ANCOVA, using the corresponding percentage of improved estimates at pretest as a covariate. This analysis yielded a significant main effect of observational learning, $F(1, 891) = 18.39$, $p < .001$, $\eta_p^2 = 0.02$, a significant main effect of information, $F(1, 891) = 103.32$, $p < .001$, $\eta_p^2 = 0.10$, and a significant interaction, $F(1, 891) = 5.21$, $p = .023$, $\eta_p^2 = 0.01$. A simple effect analysis revealed that participants in the observational learning condition improved more estimates than did participants in the control condition, $t(891) = 4.66$, $p < .001$, $d = 0.44$. However, the percentage of improved estimates was not significantly different, whether participants received the combined interventions or the information-based intervention alone, $t(891) = 1.41$, $p = .160$.

We then compared overall accuracy improvement across the four

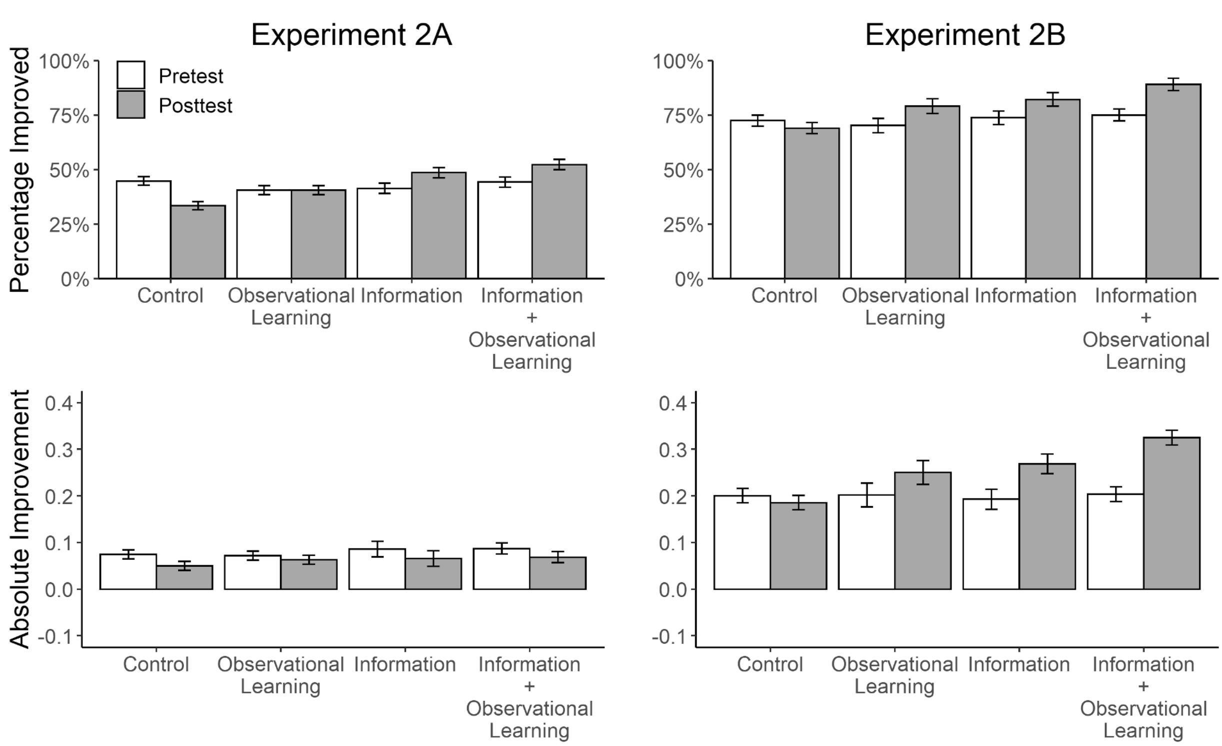

Fig. 4. Accuracy gains at pretest and posttest by training interventions, in Experiments 2A (left panel) and 2B (right panel). Top panel displays the percentage of improved estimates in each set; scale ranges from 0% to 100%; Bottom panel displays overall improvement in accuracy (i.e., estimation bias); scale ranges from negative (i.e., the estimation error is increased) to positive (i.e., the estimation error is reduced). Bars indicate 95% CIs.

conditions with a 2 (Observation: Yes vs. No) × 2 (Information: Yes vs. No) ANCOVA, using the pretest accuracy improvement as a covariate. By contrast to the percentage of revised estimates improved, the main effects of observational learning, $F(1, 891) = 2.81$, $p = .094$, and information on overall accuracy were, at best, marginal, $F(1, 891) = 3.49$, $p = .062$. There was no significant interaction, $F(1, 891) = 1.18$, $p = .278$.

Experiment 2B. We conducted parallel analyses for Experiment 2B. We first compared the percentage of revised estimates that improved at posttest across the four conditions with a 2 (Observation: Yes vs. No) × 2 (Information: Yes vs. No) ANCOVA, using the corresponding percentage of improved estimates at pretest as a covariate. This analysis yielded a significant main effect of observational learning, $F(1, 489) = 22.77$, $p < .001$, $\eta_p^2 = 0.04$, and a significant main effect of information, $F(1, 489) = 30.41$, $p < .001$, $\eta_p^2 = 0.06$. There was no significant interaction, $F(1, 489) = 1.72$, $p = .190$. A simple effect analysis revealed that participants in the observational learning condition made more improved estimates than did participants in the control condition, $t(489) = 4.33$, $p < .001$, $d = 0.55$. Indicating additivity in training effects on accuracy, participants who received the combined interventions made more improved estimates than did participants who only received the information-based training intervention, $t(489) = 2.43$, $p = .016$, $d = 0.31$.

We then compared overall accuracy improvement across the four conditions with a 2 (Observation: Yes vs. No) × 2 (Information: Yes vs. No) ANCOVA, using pretest accuracy as a covariate. This analysis yielded a significant main effect of observational learning, $F(1, 489) = 24.07$, $p < .001$, $\eta_p^2 = 0.05$, and a significant main effect of information, $F(1, 489) = 43.62$, $p < .001$, $\eta_p^2 = 0.08$. There was no significant interaction, $F(1, 489) = 0.23$, $p = .635$. A simple effect analysis revealed that participants in the observational learning condition became more accurate, overall, than did participants in the control condition, $t(489) = 3.83$, $p < .001$, $d = 0.49$. Indicating additivity in training effects on accuracy, participants who received the combined interventions also became more accurate than did participants who only received the information-based training intervention, $t(489) = 3.11$, $p = .002$, $d = 0.40$.

Together, these results provide evidence that observational learning based debiasing interventions can teach people how to make more accurate judgments by using advice. Whether their effects are additive to information-based debiasing interventions appears to depend on the quality of the advice received.

3.3. Discussion

In a WOA setting that allowed the assessment of the unique debiasing effects of observational learning, an observational learning training intervention was more effective than practice alone (i.e., the control condition) at teaching a decision rule and increasing advice taking. The effects of observational learning and an information-based training intervention were additive in their effect on learning and advice taking. Combined, the interventions more effectively taught the averaging rule and increased advice taking than the information-based intervention on its own. These effects did not vary much whether advice was naturalistic or optimized (i.e., in Experiments 2A and 2B, respectively).

The ability to determine the effects of observational learning training interventions on the accuracy of estimates was more limited in this paradigm. The optimized advice in Experiment 2B should have allowed for the detection of an effect, if one was present. The quality of naturalistic advice, however, meant that accuracy would be improved in 63% of cases if participants always adhered to the rule taught by the intervention and thus imposed a ceiling on effect detection. Our purpose in presenting this information here is to satisfy readers' natural curiosity rather than to make strong claims about improvements in accuracy due to different forms and combinations of training.

With this caveat in mind, whether advice was naturalistic or optimized, observational learning interventions tended to improve both the number of judgments improved by advice and their overall improvement in accuracy, relative to practice alone. In three out of four cases (i.e., comparing across the two accuracy dependent variables for each of the two studies), participants who received the observational learning intervention were significantly more accurate than those who received practice alone. In the fourth case the effects of observational learning marginally increased accuracy relative to practice alone.

Whether the observational learning interventions had unique debiasing effects that improved accuracy is less clear. When advice was naturalistic, observational learning interventions did not improve accuracy over and above information-based training interventions. By contrast, when advice was optimized, observational learning interventions had a complementary effect that improved accuracy. These differences in the additivity of the interventions could have been due to variation in the quality of advice provided in Experiments 2A and 2B, or to differences in the videos that participants saw. A substantive difference is the display of feedback, which may be critical for observational learning to be fully effective (Nadler et al., 2003). The intervention in Experiment 2B displayed the positive feedback received by the "actor." The intervention in Experiment 2A did not.

Importantly, the results shed light on a problem in identifying effective debiasing training interventions (Fischhoff, 1982; Milkman, Chugh, & Bazerman, 2009). The identification of effective debiasing interventions not only depends on the quality of the interventions, but also on the metrics used for their assessment. In both Experiments 2A and 2B, measures of rule learning (i.e., averaging and WOA) found observational learning interventions to have unique debiasing effects, whether the advice provided to participants was naturalistic or optimized. By contrast, measures of accuracy suggested that the interventions had unique debiasing effects only when tested in a low noise to signal environment (where advice was favorable). When advice was tested in a more naturalistic environment, there was no evidence the intervention had unique debiasing effects on accuracy. Had we chosen only one form of measurement to assess these interventions rather than both rule learning and accuracy, depending on the measure we selected, we might have come to very different conclusions about their efficacy.

4. General discussion

One-shot observational learning-based training interventions effectively reduced the commission of three consequential cognitive biases, and quickly taught a decision rule that increased advice taking. In Experiment 1, a 90-minute observational learning intervention produced a large immediate reduction in the propensity to exhibit a variety of forms of anchoring, over-reliance on the representativeness heuristic, and social projection. In Experiments 2A and 2B, 5-minute observational learning interventions immediately increased the extent to which trained participants learned the averaging principle and used advice. These findings suggest that social learning interventions, particularly those grounded in observational learning, are an effective debiasing strategy to improve judgment and decision making. We recommend social learning be added to the four strategies used in debiasing interventions to date: informing people about biases, warning them about the directionality of bias, providing feedback, and extensive coaching (Fischhoff, 1982; Milkman et al., 2009; Morewedge et al., 2015).

An important critique of debiasing training is that debiasing training interventions can impair judgment and decision making if they disrupt generally useful heuristics (Arkes, 1991; Gigerenzer & Gaissmaier, 2011). All of our experiments directly addressed this concern. In Experiment 1, all three of our bias scales included many problems which had objectively correct answers (e.g., estimates of historical facts, opinion survey data, and conjunction fallacy problems, respectively; see Appendices A–C). Observational learning interventions reduced bias and increased accuracy on these measures. The advice taking paradigm used in Experiments 2A and 2B allowed us to separate whether interventions taught a decision rule and whether they increased the accuracy of judgments. Regardless of the quality of advice that participants received in those two experiments, the results reveal that 5-minute long observational learning interventions were effective in teaching an averaging rule, and tended (in 3 out of 4 measures) to improve the accuracy of judgments relative to practice alone. The ability to identify *unique* accuracy benefits of the interventions in Experiments 2A and 2B was constrained by the quality of advice that participants incorporated in their judgments after learning the decision rule. When advice was naturalistic, an observational learning intervention modeling a decision rule did not appear to provide unique accuracy benefits above and beyond an information-based intervention that explicitly taught the decision rule. When advice was optimized, and participants saw that using the decision rule was beneficial, the observational learning and information-based training interventions appeared to be additive in their improvement of judgmental accuracy.

The additive effect of observational learning and information-based training on rule learning is particularly interesting, and worth speculating upon. One reason why this effect occurred may be the cognitive demands required for rule-based learning. Learning a rule and understanding how to implement it requires the translation of codified knowledge into a set of specific actions. Observing another person implement a rule reduces these demands (Hoover, Giambatista, & Belkin, 2012). It provides a model with which to translate that information to behavior. Second, observational learning may trigger a behavioral facilitation effect. The observed behavior can increase the accessibility of a learned rule, serving as a reminder to use it (Manz & Sims, 1981).

A third reason is that observational learning involves the acquisition of tacit knowledge, not the ability to articulate that knowledge (Nadler et al., 2003). Coupling it with information-based training may enhance its effectiveness by giving trainees the ability to explicitly articulate the principles and behaviors embedded in the tacit knowledge they have acquired. Information can facilitate the transfer of knowledge embedded in casually ambiguous tasks and organizational routines (Nelson & Winter, 1982). It may thus help overcome the problem of knowledge implementation stickiness—difficulty experienced when implementing knowledge extrapolated from a learning context—often encountered when attempting to implement learnings from observation (Szulanski, 1996, 2000; Von Hippel, 1994). Consequently, the learning process is enhanced when observation is added to information compared to information alone.

Observational learning may be a scalable and effective way to debias people, particularly in settings where it is infeasible to give direct feedback to all trainees. Our results suggest that the benefits of *hot-seating* for observers may extend to debiasing training. Hot-seating is a technique used in medicine, dentistry, and the dramatic arts, where one member of a group is tested by an interviewer or tester in front of a group (Brondani & Rossoff, 2010; Fleming, 1994; Jackson & Back, 2011; Spencer, 2003). As is the case with the teaching of medical procedures (Bong et al., 2017), the results of our experiments suggest that having trainees observe another person make bias-eliciting judgment and decisions and see the feedback they receive may be an effective debiasing training intervention.

Organizational learning research may provide insights for when observational learning will be ineffective as a debiasing training intervention. Firms seeking to learn and imitate knowledge-based innovations developed by other firms typically encounter three barriers to success: the degree to which that knowledge is complex, specific, and tacit (Kogut & Zander, 1992; McEvily & Chakravarthy, 2002). These barriers may have parallels for imitative learning between individuals. The averaging rule taught in Experiments 2A and 2B was simple. Observational learning required participants to watch and decipher what another participant did on a calculator application (i.e., add their initial estimate and the advice, and divide the total by two). More complex kinds of rules should be less easily transferred via observational learning.

Observational learning interventions may be less successful to the extent that they can only be applied to specific conditions, contexts, or problems. Whereas an averaging rule may be beneficial in a WOA paradigm where judges are relatively uncorrelated, people may continue to use it in conditions where judges are highly correlated and averaging is ineffective (Yaniv, 2004). Information may be needed to supplement observational learning when the means by which debiasing strategies improve judgments is causally ambiguous (Lippman & Rumelt, 1982; Szulanski, 1996)—when successful imitation of observed behavior requires an understanding how it produces a successful outcome. Information can help reduce such causal ambiguity, facilitating the appropriate generalization of the knowledge acquired through observational learning.

All observational learning communicates tacit knowledge, which is challenging to transfer but is often required to understand explicit knowledge (Nonaka & von Krogh, 2009). When explicit knowledge transfer is required to learn debiasing strategies and skills, observational learning may only have debiasing effects when it facilitates complementary information-based debiasing training interventions. In this light, the "combined" conditions in Experiments 2A and 2B may have achieved superior debiasing training effects because observational learning helped participants better understand how to implement the explicit knowledge embedded in the information-based training interventions.

In addition to these knowledge-based barriers, characteristics of the actor-observer relationship may modulate the effectiveness of observational learning as a debiasing intervention. In our studies, the actor was presented as a member of the group to which participants belonged (i.e., another research participant). This similarity between the actor and the observer may have amplified learning from observation (Brown & Inouye, 1978), but participants and the actor were strangers. Observational learning interventions may be even more effective if the actor and the observer trust each other and have strong social ties (Hansen, 1999; Szulanski, Cappetta, & Jensen, 2004).

5. Conclusion

Pronouncements regarding the inefficacy of debiasing training appear to have been premature (e.g., Kahneman, 2011; Milkman et al., 2009). Training is a debiasing intervention worth consideration alongside the suite of incentives and nudges used to improve judgments and decisions in business, education, law, medicine, and policy. Debiasing training interventions can require considerable effort to develop (Symborski et al., 2017). Once developed, however, they can be effectively scaled to improve decision making for many people, domains, and problems (Mellers et al., 2015; Morewedge et al., 2015). Given their scalability, low cost, and complementarity to information-based training, we believe observational learning interventions, and social learning interventions more broadly, have considerable promise to improve judgment and decision making.

Appendix A. Anchoring subscales

Item #	Subscale	Item Code	Item Description	Anchor	Correct Answer	Item Scoring
1	A	A2F0	A. Predicted happiness rating: Imagine next month that you are fired from your job. Please rate how happy you would feel during the three months following losing your job. 1 = Not at all happy; 9 = Very happy B. Revised happiness rating: At this time, please think about all of the activities you would be doing in the three months after being fired from your job. Please rate how happy you would feel during the three months following losing your job. 1 = Not at all happy; 9 = Very happy	A	N/A	|A-B|/8
2	A	A2F3	A. Predicted happiness rating: Imagine next month that you purchase and move into your dream home. Please rate how happy you would feel during the three months following moving your possessions into the new home. 1 = Not at all happy; 9 = Very happy B. Revised happiness rating: At this time, please think about all of the activities you would be doing in the three months after purchasing and moving into your dream home. Please rate how happy you would feel during the three months following moving into your dream home. 1 = Not at all happy; 9 = Very happy	A	N/A	|A-B|/8
3	A	A2F82	A. Predicted happiness rating: Imagine next month that you lose your health insurance. Please rate how happy you would feel during the three months following this event. 1 = Not at all happy; 9 = Very happy B. Revised happiness rating: At this time, please think about all of the activities you would be doing in the three months after losing your health insurance. Please rate how happy you would feel during the three months following the loss of your health insurance. 1 = Not at all happy; 9 = Very happy	A	N/A	|A-B|/8
4	A	A2F86	A. Predicted happiness rating: Imagine that you win tickets to see your favorite musical artist with backstage passes included. Please rate how happy you would feel during the three days following winning the tickets. 1 = Not at all happy; 9 = Very happy B. Revised happiness rating: At this time, please think about all of the activities you would be doing in the three days after winning tickets to see your favorite musical artist with backstage passes included. Please rate how happy you would feel during the three days following winning the tickets. 1 = Not at all happy; 9 = Very happy	A	N/A	|A-B|/8
5	A	A1cER2a3	Think about whether “The Godfather” first appeared in theaters before or after 1991. When did “The Godfather” first appear in theaters?	1991	1972	|Correct Answer – Answer| / |Correct Answer – Anchor|
6	A	A1cER3a3	Think about whether James K. Polk began his term as U.S. president before or after 1875.When did James K. Polk begin his term as U.S. president?	1875	1845	|Correct Answer – Answer| / |Correct Answer – Anchor|
7	A	A1cER11a2	Think about whether the average temperature in New York in September is higher or lower than 58 degrees Fahrenheit.What is the average temperature (in degrees Fahrenheit) in New York in September?	58	74	|Correct Answer – Answer| / |Correct Answer – Anchor|
8	A	A1cER38a2	The American Civil war began when the southern states seceded from the United States of America. Only about 1,700 American soldiers were killed during the Revolutionary War. About how many American soldiers were killed in total during the Civil War?	1700	750,000	|Correct Answer – Answer| / |Correct Answer – Anchor|
9	A	A1cER84a3	Think about whether “Taxi Driver” first appeared in movie theaters before or after 1988. When did “Taxi Driver” first appear in theaters?	1988	1976	

(*continued on next page*)

(*continued*)

Item #	Subscale	Item Code	Item Description	Anchor	Correct Answer	Item Scoring
						|Correct Answer – Answer| / |Correct Answer – Anchor|
10	A	A1cER107a3	Think about whether the state of Rhode Island is wider than 200 miles. How wide (in miles) is the state of Rhode Island?	200	37	|Correct Answer – Answer| / |Correct Answer – Anchor|
11	A	A1cER160a2	Think about whether the Mississippi River is shorter or longer than 500 miles. What is your estimate of the exact length of the Mississippi River (in miles)?	500	2320	|Correct Answer – Answer| / |Correct Answer – Anchor|
12	A	A1dSR2	In what year did the Boston Tea Party occur?	1776	1773	|Correct Answer – Answer| / |Correct Answer – Anchor|
13	A	A1dSR7	What is the gestation period (in months) of a puma?	9	3	|Correct Answer – Answer| / |Correct Answer – Anchor|
14	A	A1aEI0a2	A number between 0 and 100 is determined by spinning a wheel of fortune. The number lands of 41. Think about whether the percentage of African nations that have national soccer teams is higher or lower than 41%. What is the percentage of African nations that have national soccer teams?	41	95	|Correct Answer – Answer| / |Correct Answer – Anchor|
15	A	A1aEI87	Think about whether a South American sloth has more or less than 4 toes. What is the maximum speed of a NASCAR during a race (in miles-per-hour)?	4	215	|Correct Answer – Answer| / |Correct Answer – Anchor|
16	A	A1aEI92a2	Think about whether the tallest recorded human is taller or shorter than 10 feet. How much does an average pineapple weigh (in pounds)?	10	1.5	|Correct Answer – Answer| / |Correct Answer – Anchor|
17	A	A1bSI4	A. How many total eyelets do the shoes you are wearing have in them (the holes that shoelaces go through)?B. Think about whether the average low January temperature (in degrees Fahrenheit) in St. Louis is above or below the number you gave in the previous question. What is the average low January temperature (in degrees Fahrenheit) in St. Louis?	A	24	|Correct Answer – Answer| / |Correct Answer – Anchor|
18	A	A1bSI34	A. Create a four digit number using the last four digits of your social security number. Enter that number below.B. Think about whether Van Gogh painted The Starry Night before or after this date (use the number you gave in the previous question).In what year did Van Gogh paint The Starry Night?	A	1889	|Correct Answer – Answer| / |Correct Answer – Anchor|
19	B	A2F1	A. Predicted happiness rating: Imagine next month that you get a job and move to Emporia, Kansas, a rural town 90 miles from Wichita. Please rate how happy you would feel during the three months following your arrival in Emporia. 1 = Not at all happy; 9 = Very happy B. Revised happiness rating: At this time, please think about all of the activities you would be doing in the three months after moving to Emporia, Kansas. Please rate how happy you would feel during the three months following your arrival in Emporia. 1 = Not at all happy; 9 = Very happy	A		|A-B|/8
20	B	A2F4	A. Predicted happiness rating: Imagine next month that you win the lottery. Please rate how happy you would feel during the three months following this event. 1 = Not at all happy; 9 = Very happy B. Revised happiness rating: At this time, please think about all of the activities you would be doing in the three months after winning the lottery.	A		|A-B|/8

(*continued on next page*)

(*continued*)

Item #	Subscale	Item Code	Item Description	Anchor	Correct Answer	Item Scoring
			Please rate how happy you would feel during the three months following winning the lottery. 1 = Not at all happy; 9 = Very happy			
21	B	A2F5	A. Predicted happiness rating: Imagine that you invest a substantial amount of your savings in a particular stock and it crashes, losing all of your money. Please rate how happy you would feel during the first three months after the stock crash. 1 = Not at all happy; 9 = Very happy At this time, please think about all of the activities you would be doing in the three months after investing a substantial amount of your savings in a particular stock and it crashes, losing all of your money. Please rate how happy you would feel during the three months following the stock crash. 1 = Not at all happy; 9 = Very happy	A		|A-B|/8
22	B	A2F81	A. Predicted happiness rating: Imagine next month that you get a substantial pay raise in your current job. Please rate how happy you would feel during the three months following this event. 1 = Not at all happy; 9 = Very happy B. Revised happiness rating: At this time, please think about all of the activities you would be doing in the three months after getting a substantial pay raise in your current job. Please rate how happy you would feel during the three months following your pay raise. 1 = Not at all happy; 9 = Very happy	A		|A-B|/8
23	B	A1cER21a3	Think about whether the average winter temperature in Antarctica is higher or lower than 1 degree Fahrenheit. What is the average winter temperature in Antarctica (in degrees Fahrenheit)?	1	−30	|Correct Answer – Answer| / |Correct Answer – Anchor|
24	B	A1cER27a3	Think about whether the tip of the Empire State Building is greater or less than 2465 feet high. How high (in feet) is the tip of the Empire State Building?	2465	1454	|Correct Answer – Answer| / |Correct Answer – Anchor|
25	B	A1cER35a3	Think about whether the electric vacuum was invented before or after 1950. When was the electric vacuum invented?	1950	1907	|Correct Answer – Answer| / |Correct Answer – Anchor|
26	B	A1cER83a3	Think about whether “Casablanca” first appeared in theaters before or after 1969. When did “Casablanca” first appear in theaters?	1969	1942	|Correct Answer – Answer| / |Correct Answer – Anchor|
27	B	A1cER89a2	Think about whether actress Meryl Streep is older or younger than 47 years. What is the age of actress Meryl Streep?	47	64	|Correct Answer – Answer| / |Correct Answer – Anchor|
28	B	A1cER103a3	Think about whether the height of the tallest redwood tree is greater or less than 1516 feet. How tall (in feet) is the height of the tallest redwood tree?	1516	367	|Correct Answer – Answer| / |Correct Answer – Anchor|
29	B	A1cER104a2	Think about whether the number of nations in the United Nations is greater or fewer than 20. What is the number of nations in the United Nations?	20	193	|Correct Answer – Answer| / |Correct Answer – Anchor|
30	B	A1dSR1	How many states were in the United States in 1880?	50	38	|Correct Answer – Answer| / |Correct Answer – Anchor|
31	B	A1dSR6	What is the freezing point (in degrees Fahrenheit) of saltwater?	32	28	|Correct Answer – Answer| / |Correct Answer – Anchor|
32	B	A1aEI2a2	Giraffes can run about 8 miles per hour. How long is a giraffe’s tongue (in inches)?	8	20	

(*continued on next page*)

(*continued*)

Item #	Subscale	Item Code	Item Description	Anchor	Correct Answer	Item Scoring
						|Correct Answer – Answer| / |Correct Answer – Anchor|
33	B	A1aEI7a3	Think about whether the monthly rent of an average one bedroom apartment in Pittsburgh is more or less than $2000. In what year did the Spanish Inquisition take place?	2000	1478	|Correct Answer – Answer| / |Correct Answer – Anchor|
34	B	A1aEI89	Think about whether a 1 lb coconut is above or below the maximum weight that a swallow can carry. What is the gestation period of an African Elephant (in months)?	1	22	|Correct Answer – Answer| / |Correct Answer – Anchor|
35	B	A1bSI8	A. What are the last four numbers in your phone number?B. Now treat the number you just generated as a date (year) and think about whether the French Revolution began before or after this year. In what year did the French Revolution start?	A	1789	|Correct Answer – Answer| / |Correct Answer – Anchor|
36	B	A1bSI32	A. About how many days remain until your next birthday?B. Think about whether the average high temperature in July in Death Valley, CA is above or below the number you gave in the previous question. What is the average high temperature (in degrees Fahrenheit) in July in Death Valley, CA?	A	116	|Correct Answer – Answer| / |Correct Answer – Anchor|
37	C	A2F2	A. Predicted happiness rating: Imagine next month that you get into an accident and become paraplegic. Please rate how happy you would feel during the three months following this event. 1 = Not at all happy; 9 = Very happy B. Revised happiness rating: At this time, please think about all of the activities you would be doing in the three months after getting into an accident and becoming paraplegic. Please rate how happy you would feel during the three months after becoming paraplegic. 1 = Not at all happy; 9 = Very happy	A		|A-B|/8
38	C	A2F80	A. Predicted happiness rating: Imagine next month that you get a job and move to California. Please rate how happy you would feel during the three months following your arrival in California. 1 = Not at all happy; 9 = Very happy B. Revised happiness rating: At this time, please think about all of the activities you would be doing in the three months after arriving in California. Please rate how happy you would feel during the three months following your arrival in California. 1 = Not at all happy; 9 = Very happy	A		|A-B|/8
39	C	A2F83	A. Predicted happiness rating: Imagine next month that you receive excellent health benefits (substantially better than your current ones). Please rate how happy you would feel during the three months following this event. 1 = Not at all happy; 9 = Very happy B. Revised happiness rating: At this time, please think about all of the activities you would be doing in the three months after being fired from your job. Please rate how happy you would feel during the three months following losing your job. 1 = Not at all happy; 9 = Very happy	A		|A-B|/8
40	C	A2F85	A. Predicted happiness rating: Imagine that you lose your wallet (containing a significant amount of cash, your credit cards, and driver's license). Please rate how happy you would feel during the three days following this event. 1 = Not at all happy; 9 = Very happyB. Revised happiness rating: At this time, please think about all of the activities you would be doing in the three days after losing your wallet (containing a significant amount of cash, your credit cards, and driver's license) . Please rate how happy you would feel during the three days following losing your wallet. 1 = Not at all happy; 9 = Very happy	A		|A-B|/8
41	C	A1cER7a3	Think about whether Jennifer Lopez's birth year is before or after 1988. What is Jennifer Lopez's birth year?	1988	1970	|Correct Answer – Answer| / |Correct Answer – Anchor|
42	C	A1cER30a2		25	53	

(*continued on next page*)

(*continued*)

Item #	Subscale	Item Code	Item Description	Anchor	Correct Answer	Item Scoring
			Think about whether, on average, >25 people are killed in the United States by lightning strikes each year. How many people, on average, are killed in the United States by lightning strikes each year?			|Correct Answer – Answer| / |Correct Answer – Anchor|
43	C	A1cER34a3	Think about whether the microwave was invented before or after 1974. When was the microwave invented?	1974	1945	|Correct Answer – Answer| / |Correct Answer – Anchor|
44	C	A1cER90a2	Think about whether the average annual rainfall in Philadelphia is greater or less than 25 in.. What is the average annual rainfall in Philadelphia (in inches)?	25	41	|Correct Answer – Answer| / |Correct Answer – Anchor|
45	C	A1cER93a2	Think about whether the average temperature in Montreal in December is higher or lower than 3 degrees Fahrenheit. What is the average temperature in Montreal in December (in degrees Fahrenheit)?	3	26	|Correct Answer – Answer| / |Correct Answer – Anchor|
46	C	A1cER95a2	Think about whether the maximum speed of a housecat is greater or less than 5 miles per hour. What is the maximum speed of a housecat (in miles-per-hour)?	5	30	|Correct Answer – Answer| / |Correct Answer – Anchor|
47	C	A1cER97a3	Consider whether the number of interstate highways is more or less than 500. How many interstate highways are there in the continental United States?	500	66	|Correct Answer – Answer| / |Correct Answer – Anchor|
48	C	A1dSR5	What is the duration (in days) of Mars' orbit around the Sun?	365	687	|Correct Answer – Answer| / |Correct Answer – Anchor|
49	C	A1dSR10	In what year did the second European explorer, after Columbus, land in the West Indies?	1492	1501	|Correct Answer – Answer| / |Correct Answer – Anchor|
50	C	A1aEI5a3	Think about whether a standard cricket field is longer or shorter than 100 feet. What percentage of the US population is Catholic?	100	25	|Correct Answer – Answer| / |Correct Answer – Anchor|
51	C	A1aEI6a3	Think about whether there are more or less than 600 people living in a square block in Washington D.C. How many miles separate Washington, D. C., and New York City?	600	225	|Correct Answer – Answer| / |Correct Answer – Anchor|
52	C	A1aEI91a2	Think about whether the average price of a home in Beverly Hills is more or less than $5,000,000. What is the population of Sacramento, CA?	5,000,000	475,000	|Correct Answer – Answer| / |Correct Answer – Anchor|
53	C	A1bSI6	A. How tall are you (in inches)?B. Think about whether the average price (in US dollars) of a sofa bed at IKEA is above or below the number you gave in the previous question. What is the average price (in US dollars) of a sofa bed at IKEA?	A	693	|Correct Answer – Answer| / |Correct Answer – Anchor|
54	C	A1bSI36	A. How old are you?B. Think about whether this number (use the number you gave in the previous question) is more or less than the number of novels Stephen King has written. How many novels has Stephen King written?	A	51	|Correct Answer – Answer| / |Correct Answer – Anchor|

*For anchoring items, estimates that fell outside the range between the anchor and the correct answer were coded as 1, if the estimate was in the direction opposite to the anchor or as 0, if the estimate was in the same direction as the anchor.

Appendix B. Projection subscales

Item #	Subscale	Item #	Item Description	Actual Consensus	Item Scoring
1	A	Pro1b_FalseConFx_13	A. Is it a good thing that Supreme Court justices get lifetime appointments? (Yes/No)B. What percentage of Americans do you think agrees with your response? (0–100%)	Yes: 33% No: 60%	1 – (%B. – %actual) / (100 - %actual)
2	A	Pro1b_FalseConFx_16	A. Do you think the United States has a responsibility to do something about the fighting in Syria between government forces and anti-government groups? (Yes/No)B. What percentage of Americans do you think agrees with your response? (0–100%)	Yes: 24% No: 62%	1 – (%B. – %actual) / (100 - %actual)
3	A	Pro1b_FalseConFx_18	A. Do you think the U.S. should take the leading role among all other countries in the world in trying to solve international conflicts? (Yes/No)B. What percentage of Americans do you think agrees with your response? (0–100%)	Yes: 35% No: 52%	1 – (%B. – %actual) / (100 - %actual)
4	A	Pro1b_FalseConFx_19	A. Cyber-attacks are increasingly being used as a new tactic of warfare by some countries. Do you think the U.S. should ever conduct cyber-attacks against other countries? (Yes/No)B. What percentage of Americans do you think agrees with your response? (0–100%)	Yes: 36% No: 55%	1 – (%B. – %actual) / (100 - %actual)
5	A	Pro1b_FalseConFx_20	A. Would you favor or oppose the U.S. taking military action against Iran in order to prevent them from producing a nuclear weapon? (Favor/Oppose)B. What percentage of Americans do you think agrees with your response? (0–100%)	Favor: 58% Oppose: 37%	1 – (%B. – %actual) / (100 - %actual)
6	A	Pro1b_FalseConFx_24	A. Do you think that gambling is morally acceptable or morally wrong? (Morally acceptable/ Morally wrong)B. What percentage of Americans do you think agrees with your response? (0–100%)	Morally acceptable: 64% Morally wrong: 31%	1 – (%B. – %actual) / (100 - %actual)
7	A	Pro1b_FalseConFx_29	A. Do you think that buying and wearing clothing made of animal fur is morally acceptable or morally wrong? (Morally acceptable/ Morally wrong)B. What percentage of Americans do you think agrees with your response? (0–100%)	Morally acceptable: 59% Morally wrong: 36%	1 – (%B. – %actual) / (100 - %actual)
8	A	Pro1b_FalseConFx_37	A. Do you think that suicide is morally acceptable or morally wrong? (Morally acceptable/ Morally wrong)B. What percentage of Americans do you think agrees with your response? (0–100%)	Morally acceptable: 16% Morally wrong: 77%	1 – (%B. – %actual) / (100 - %actual)
9	A	Pro1b_FalseConFx_44	A. Do you think that state governments should provide more education assistance? (Yes/No)B. What percentage of Americans do you think agrees with your response? (0–100%)	Morally acceptable: 59% Morally wrong: 41%	1 – (%B. – %actual) / (100 - %actual)
10	A	Pro1b_FalseConFx_51	A. Do you think that it is easier for a married person than a single person to raise a family? (Yes/No)B. What percentage of Americans do you think agrees with your response? (0–100%)	Yes: 77% No: 23%	1 – (%B. – %actual) / (100 - %actual)
11	A	Pro1b_FalseConFx_54	A. Do you drink coffee? (Yes/No)B. What percentage of Americans (over 18) do you think drinks coffee? (0–100%)	Yes: 54% No: 46%	1 – ((100 –%B.) – %actual) / (100 - %actual)
12	A	Pro1b_FalseConFx_69	A. Have you ever searched online for medical information (i.e., to diagnose a problem you were having)? (Yes/No) B. What percentage of adult Americans do you think have ever searched online for medical information (i.e., to diagnose a problem they were having)? (0–100%)	Yes: 35% No: 65%	1 – ((100 –%B.) – %actual) / (100 - %actual)
13	A	Pro1b_FalseConFx_84	A. Would you say, on average, you drink at least one soda a day? (Yes/No)B. What percentage of Americans do you think drinks at least one soda a day? (0–100%)	Yes: 48% No: 52%	1 – ((100 –%B.) – %actual) / (100 - %actual)
14	A	Pro2a_AttSim_0	I think (The average student thinks) it is more fun to be involved in a discussion where there is a lot of disagreement. (1 = strongly disagree; 2 = disagree; 3 = neutral; 4 = agree; 5 = strongly agree)		∣ self-rating – other-rating ∣ / ∣ max possible difference score based on self-rating ∣
15	A	Pro2a_AttSim_1	I (the average student) enjoy(s) talking to people with points of view different than mine (his or hers). (1 = strongly disagree; 2 = disagree; 3 = neutral; 4 = agree; 5 = strongly agree)		∣ self-rating – other-rating ∣ / ∣ max possible difference score based on self-rating ∣
16	A	Pro2a_AttSim_7	I (the average student) enjoy(s) arguing with other people about things on which we (they) disagree. (1 = strongly disagree; 2 = disagree; 3 = neutral; 4 = agree; 5 = strongly agree)		∣ self-rating – other-rating ∣ / ∣ max possible difference score based on self-rating ∣
17	A	Pro2a_AttSim_8	I (the average student) would prefer to work independently rather than to work with other people and have disagreements. (1 = strongly disagree; 2 = disagree; 3 = neutral; 4 = agree; 5 = strongly agree)		∣ self-rating – other-rating ∣ / ∣ max possible difference score based on self-rating ∣

(*continued on next page*)

(*continued*)

Item #	Subscale	Item #	Item Description	Actual Consensus	Item Scoring
18	A	Pro2a_AttSim_13	I (the average student) enjoy(s) disagreeing with others. (1 = strongly disagree; 2 = disagree; 3 = neutral; 4 = agree; 5 = strongly agree)		| self-rating – other-rating | / | max possible difference score based on self-rating |
19	A	Pro2c_AttSim_1	A. To what extent do you agree with the following statement?I am energetic. (1 = not at all; 7 = definitely)B. Imagine you meet another study participant. Do you think that he or she is likely to be energetic? (1 = not at all; 7 = definitely)		| self-rating – other-rating | / | max possible difference score based on self-rating |
20	A	Pro2c_AttSim_2	A. To what extent do you agree with the following statement?I am adventurous. (1 = not at all; 7 = definitely)B. Imagine you meet another study participant. Do you think that he or she is likely to be adventurous? (1 = not at all; 7 = definitely)		| self-rating – other-rating | / | max possible difference score based on self-rating |
21	A	Pro2c_AttSim_4	A. To what extent do you agree with the following statement?I am outgoing. (1 = not at all; 7 = definitely)B. Imagine you meet another study participant. Do you think that he or she is likely to be outgoing? (1 = not at all; 7 = definitely)		| self-rating – other-rating | / | max possible difference score based on self-rating |
22	A	Pro2c_AttSim_5	A. To what extent do you agree with the following statement?I am forgiving. (1 = not at all; 7 = definitely)B. Imagine you meet another study participant. Do you think that he or she is likely to be forgiving? (1 = not at all; 7 = definitely)		| self-rating – other-rating | / | max possible difference score based on self-rating |
23	A	Pro2c_AttSim_18	A. To what extent do you agree with the following statement?I am irritable. (1 = not at all; 7 = definitely)B. Imagine you meet another study participant. Do you think that he or she is likely to be irritable? (1 = not at all; 7 = definitely)		| self-rating – other-rating | / | max possible difference score based on self-rating |
24	B	Pro1b_FalseConFx_0	A. Do you approve or disapprove of the way Barack Obama is handling foreign policy?B. What percentage of Americans do you think agrees with your response? (0–100%)	Approve: 41% Disapprove: 52%	1 – (%B. – %actual) / (100 - %actual)
25	B	Pro1b_FalseConFx_4	A. Do you approve or disapprove of the way Barack Obama is handling gun policy?B. What percentage of Americans do you think agrees with your response? (0–100%)	Approve: 41% Disapprove: 52%	1 – (%B. – %actual) / (100 - %actual)
26	B	Pro1b_FalseConFx_14	A. If North Korea were to attack South Korea, should the U.S. use its military forces to help defend South Korea? (Yes/No)B. What percentage of Americans do you think agrees with your response? (0–100%)	Yes: 55% No: 34%	1 – (%B. – %actual) / (100 - %actual)
27	B	Pro1b_FalseConFx_34	A. Do you think that cloning animals is morally acceptable or morally wrong? (Morally acceptable/ Morally wrong)B. What percentage of Americans do you think agrees with your response? (0–100%)	Morally acceptable: 34% Morally wrong: 60%	1 – (%B. – %actual) / (100 - %actual)
28	B	Pro1b_FalseConFx_42	A. Do you think that higher education institutions should reduce tuition and fees? (Yes/No)B. What percentage of Americans do you think agrees with your response? (0–100%)	Yes: 67% No: 33%	1 – (%B. – %actual) / (100 - %actual)
29	B	Pro1b_FalseConFx_53	A. Do you think that it should be legal for same-sex couples to marry? (Yes/No)B. What percentage of Americans do you think agrees with your response? (0–100%)	Yes: 53% No: 39%	1 – (%B. – %actual) / (100 - %actual)
30	B	Pro1b_FalseConFx_55	A. Do you own a smartphone? (Yes/No)B. What percentage of Americans do you think owns a smartphone? (0–100%)	Yes: 56% No: 44%	1 – ((100 –%B.) – %actual) / (100 - %actual)
31	B	Pro1b_FalseConFx_57	A. Do you have magnets on your refrigerator? (Yes/No)B. What percentage of Americans do you think has magnets on their refrigerator? (0–100%)	Yes: 87% No: 12%	1 – ((100 –%B.) – %actual) / (100 - %actual)
32	B	Pro1b_FalseConFx_65	A. Do you currently use Facebook? (Yes/No)B. What percentage of Americans do you think currently uses Facebook? (0–100%)	Yes: 67% No: 33%	1 – ((100 –%B.) – %actual) / (100 - %actual)
33	B	Pro1b_FalseConFx_68	A. Have you searched online for medical information (i.e., to diagnose a problem you were having) in the past year? (Yes/ No) B. What percentage of adult Americans do you think has searched online for medical information (i.e., to diagnose a problem they were having) in the past year? (0–100%)	Yes: 59% No: 41%	1 – ((100 –%B.) – %actual) / (100 - %actual)
34	B	Pro1b_FalseConFx_81	A. Which do you like better: dogs or cats? (Dogs/Cats)B. What percentage of Americans do you think agrees with your opinion? (0–100%)	Dogs: 52% Cats: 21%	1 – (%B. – %actual) / (100 - %actual)
35	B	Pro1b_FalseConFx_82	A. Do you believe the Loch Ness Monster is real? (Real/Not real)B. What percentage of Americans do you think agrees with your opinion? (0–100%)	Real: 18% Not real: 64%	1 – (%B. – %actual) / (100 - %actual)
36	B	Pro1b_FalseConFx_92		Yes: 53% No: 6%	1 – (%B. – %actual) / (100 - %actual)

(*continued on next page*)

(*continued*)

Item #	Subscale	Item #	Item Description	Actual Consensus	Item Scoring
			A. Have you ever watched a television show by streaming it over the internet? (Yes/No)B. What percentage of Americans do you think also have/have not watched a television show by streaming it over the internet? (0–100%)		
37	B	Pro2a_AttSim_2	I (the average student doesn't) don't like to be in situations where people are in disagreement. (1 = strongly disagree; 2 = disagree; 3 = neutral; 4 = agree; 5 = strongly agree)		| self-rating – other-rating | / | max possible difference score based on self-rating |
38	B	Pro2a_AttSim_3	I (the average student) prefer(s) being in groups where everyone's beliefs are the same as mine (his/hers). (1 = strongly disagree; 2 = disagree; 3 = neutral; 4 = agree; 5 = strongly agree)		| self-rating – other-rating | / | max possible difference score based on self-rating |
39	B	Pro2a_AttSim_5	I (the average student) prefer(s) to change the topic of discussion when disagreement occurs. (1 = strongly disagree; 2 = disagree; 3 = neutral; 4 = agree; 5 = strongly agree)		| self-rating – other-rating | / | max possible difference score based on self-rating |
40	B	Pro2a_AttSim_6	I (the average student) tend(s) to create disagreements in conversations because it serves a useful purpose. (1 = strongly disagree; 2 = disagree; 3 = neutral; 4 = agree; 5 = strongly agree)		| self-rating – other-rating | / | max possible difference score based on self-rating |
41	B	Pro2a_AttSim_14	Disagreement stimulates a conversation and causes me (the average student) to communicate more. (1 = strongly disagree; 2 = disagree; 3 = neutral; 4 = agree; 5 = strongly agree)		| self-rating – other-rating | / | max possible difference score based on self-rating |
42	B	Pro2c_AttSim_3	A. To what extent do you agree with the following statement?I am enthusiastic. (1 = not at all; 7 = definitely)B. Imagine you meet another study participant. Do you think that he or she is likely to be enthusiastic? (1 = not at all; 7 = definitely)		| self-rating – other-rating | / | max possible difference score based on self-rating |
43	B	Pro2c_AttSim_7	A. To what extent do you agree with the following statement?I am warm. (1 = not at all; 7 = definitely)B. Imagine you meet another study participant. Do you think that he or she is likely to be warm? (1 = not at all; 7 = definitely)		| self-rating – other-rating | / | max possible difference score based on self-rating |
44	B	Pro2c_AttSim_17	A. To what extent do you agree with the following statement?I am tense. (1 = not at all; 7 = definitely)B. Imagine you meet another study participant. Do you think that he or she is likely to be tense? (1 = not at all; 7 = definitely)		| self-rating – other-rating | / | max possible difference score based on self-rating |
45	B	Pro2c_AttSim_22	A. To what extent do you agree with the following statement?I am not self-confident. (1 = not at all; 7 = definitely)B. Imagine you meet another study participant. Do you think that he or she is likely to be not self-confident? (1 = not at all; 7 = definitely)		| self-rating – other-rating | / | max possible difference score based on self-rating |
46	B	Pro2c_AttSim_27	A. To what extent do you agree with the following statement?I am excitable. (1 = not at all; 7 = definitely)B. Imagine you meet another study participant. Do you think that he or she is likely to be excitable? (1 = not at all; 7 = definitely)		| self-rating – other-rating | / | max possible difference score based on self-rating |
47	C	Pro1b_FalseConFx_1	A. Do you approve or disapprove of the way Barack Obama is handling the economy? (Approve/ Disapprove)B. What percentage of Americans do you think agrees with your response? (0–100%)	Approve: 45% Disapprove: 38%	1 – (%B. – %actual) / (100 - %actual)
48	C	Pro1b_FalseConFx_2	A. Do you approve or disapprove of the way Barack Obama is handling the federal budget deficit? (Approve/ Disapprove)B. What percentage of Americans do you think agrees with your response? (0–100%)	Approve: 33% Disapprove: 57%	1 – (%B. – %actual) / (100 - %actual)
49	C	Pro1b_FalseConFx_3	A. Do you approve or disapprove of the way Barack Obama is handling immigration? (Approve/ Disapprove)B. What percentage of Americans do you think agrees with your response? (0–100%)	Approve: 44% Disapprove: 45%	1 – (%B. – %actual) / (100 - %actual)
50	C	Pro1b_FalseConFx_8	A. Do you think that upper-income Americans pay too little in federal taxes? (Yes/No)B. What percentage of Americans do you think agrees with your response? (0–100%)	Yes: 58% No: 42%	1 – (%B. – %actual) / (100 - %actual)
51	C	Pro1b_FalseConFx_25	A. Do you think that sex between an unmarried man and woman is morally acceptable or morally wrong? (Morally acceptable/ Morally wrong)B. What percentage of Americans do you think agrees with your response? (0–100%)	Morally acceptable: 63% Morally wrong: 33%	1 – (%B. – %actual) / (100 - %actual)
52	C	Pro1b_FalseConFx_31	A. Do you think that medical testing on animals is morally acceptable or morally wrong? (Morally acceptable/ Morally wrong)B. What percentage of Americans do you think agrees with your response? (0–100%)	Morally acceptable: 56% Morally wrong: 39%	1 – (%B. – %actual) / (100 - %actual)
53	C	Pro1b_FalseConFx_35	A. Do you think that sex between teenagers is morally acceptable or morally wrong? (Morally acceptable/ Morally wrong) B. What percentage of Americans do you think agrees with your response? (0–100%)	Morally acceptable: 32% Morally wrong: 63%	1 – (%B. – %actual) / (100 - %actual)
54	C	Pro1b_FalseConFx_43	A. Do you think that the higher education system in the U.S. fails to provide students with good value for the money that they and their families spend? (Yes/No)B. What percentage of Americans do you think agrees with your response? (0–100%)	Yes: 57% No: 43%	1 – (%B. – %actual) / (100 - %actual)

(*continued on next page*)

(*continued*)

Item #	Subscale	Item #	Item Description	Actual Consensus	Item Scoring
55	C	Pro1b_FalseConFx_45	A. Do you think that the federal government should provide more education assistance? (Yes/No)B. What percentage of Americans do you think agrees with your response? (0–100%)	Yes: 55% No: 45%	1 – (%B. – %actual) / (100 - %actual)
56	C	Pro1b_FalseConFx_47	A. Do you think that the cost of higher education is affordable to anyone who needs it? (Yes/No)B. What percentage of Americans do you think agrees with your response? (0–100%)	Yes: 26% No: 74%	1 – (%B. – %actual) / (100 - %actual)
57	C	Pro1b_FalseConFx_50	A. Do you think that having a successful marriage is "one of the most important things in life"? (Yes/No)B. What percentage of Americans do you think agrees with your response? (0–100%)	Yes: 36% No: 64%	1 – (%B. – %actual) / (100 - %actual)
58	C	Pro1b_FalseConFx_70	A. Have you ever paid for online content (e.g., music, software, games)? (Yes/No) B. What percentage of adult Americans do you think have ever paid for online content (e.g., music, software, games)? (0–100%)	Yes: 65% No: 35%	1 – ((100 –%B.) – %actual) / (100 - %actual)
59	C	Pro1b_FalseConFx_78	A. Do you believe the Bush administration intentionally misled the public about the possibility of weapons of mass destruction in Iraq to promote the Iraq War? (Yes/No)B. What percentage of Americans do you think believe the Bush administration intentionally misled the public about the possibility of weapons of mass destruction in Iraq to promote the Iraq War? (0–100%)	Yes: 44% No: 45%	1 – ((100 –%B.) – %actual) / (100 - %actual)
60	C	Pro2a_AttSim_4	I believe (the average student believes) disagreements are generally personalized. (1 = strongly disagree; 2 = disagree; 3 = neutral; 4 = agree; 5 = strongly agree)		| self-rating – other-rating | / | max possible difference score based on self-rating |
61	C	Pro2a_AttSim_9	I (the average student) would prefer joining a group where no disagreements occur. (1 = strongly disagree; 2 = disagree; 3 = neutral; 4 = agree; 5 = strongly agree)		| self-rating – other-rating | / | max possible difference score based on self-rating |
62	C	Pro2a_AttSim_10	I don't (the average student doesn't) like to disagree with other people. (1 = strongly disagree; 2 = disagree; 3 = neutral; 4 = agree; 5 = strongly agree)		| self-rating – other-rating | / | max possible difference score based on self-rating |
63	C	Pro2a_AttSim_11	Given a choice, I (the average student) would leave a conversation rather than continue a disagreement. (1 = strongly disagree; 2 = disagree; 3 = neutral; 4 = agree; 5 = strongly agree)		| self-rating – other-rating | / | max possible difference score based on self-rating |
64	C	Pro2a_AttSim_12	I (the average student) avoid(s) talking with people who I think will disagree with me. (1 = strongly disagree; 2 = disagree; 3 = neutral; 4 = agree; 5 = strongly agree)		| self-rating – other-rating | / | max possible difference score based on self-rating |
65	C	Pro2c_AttSim_6	A. To what extent do you agree with the following statement?I am not demanding. (1 = not at all; 7 = definitely)B. Imagine you meet another study participant. Do you think that he or she is likely to be not demanding? (1 = not at all; 7 = definitely)		| self-rating – other-rating | / | max possible difference score based on self-rating |
66	C	Pro2c_AttSim_11	A. To what extent do you agree with the following statement?I am efficient. (1 = not at all; 7 = definitely)B. Imagine you meet another study participant. Do you think that he or she is likely to be efficient? (1 = not at all; 7 = definitely)		| self-rating – other-rating | / | max possible difference score based on self-rating |
67	C	Pro2c_AttSim_14	A. To what extent do you agree with the following statement?I am thorough. (1 = not at all; 7 = definitely)B. Imagine you meet another study participant. Do you think that he or she is likely to be thorough? (1 = not at all; 7 = definitely)		| self-rating – other-rating | / | max possible difference score based on self-rating |
68	C	Pro2c_AttSim_20	A. To what extent do you agree with the following statement?I am shy. (1 = not at all; 7 = definitely)B. Imagine you meet another study participant. Do you think that he or she is likely to be shy? (1 = not at all; 7 = definitely)		| self-rating – other-rating | / | max possible difference score based on self-rating |
69	C	Pro2c_AttSim_21	A. To what extent do you agree with the following statement?I am moody. (1 = not at all; 7 = definitely)B. Imagine you meet another study participant. Do you think that he or she is likely to be moody? (1 = not at all; 7 = definitely)		| self-rating – other-rating | / | max possible difference score based on self-rating |

Appendix C. Representativeness subscales

Item	Subscale	Item Code	Item	Type	Answer
1	A	Rep1a.07	You are at a coffee shop and you see a man sitting on the patio, smoking a cigarette. The man has arms covered in tattoos, multiple visible piercings, and his hair is styled in a mohawk. The man is reading a book, and taking notes on a laptop. Based on this information, please rank the following in terms of probability (1 = highest probability; 4 = lowest probability): A. The man is a tattoo artist. B. The man smokes cigarettes.C. The man spends his weekends at rock concerts.D. The man is a tattoo artist who spends his weekends at rock concerts.	Conjunction Fallacy	D should be ranked lower than A and C
2	A	Rep1e.02	Scott loves plants, being outdoors, and built a large community garden for a service project in high school. Which of the following is most likely? A. Scott is an IT technician B. Scott is an IT technician who grows herbs at home for cooking	Conjunction Fallacy	A
3	A	Rep1e.03	Daniel is a gifted individual with above average IQ. Socially awkward, Daniel refuses to shake hands because of germs. His great memory skills have allowed him to memorize the capital of every country in the world. Daniel has a strong desire to succeed and does not exhibit grace in defeat. Daniel is not particularly athletic and does not enjoy exercise. Daniel was often made fun of in school for being a “nerd.” Which of the following is most likely? A. Daniel is a janitor B. Daniel is a janitor who plays chess on the weekends	Conjunction Fallacy	A
4	A	Rep1e.09	Roger is 52 years old, lives alone, and shows signs of early aging. He spends most of his time in his apartment sipping tea, typing on his iPad, or reading an eBook. When he does show his face in public, he is very soft spoken but loves to tell stories to children. Which is most likely? A. Roger is an engineer. B. Roger is an engineer who is also an avid writer on the side.	Conjunction Fallacy	A
5	A	Rep1f.01	Peter is reading bios on an on-line dating website. He comes across Janet’s bio and reads that she is quiet, loves to read books, do crossword puzzles, and was awarded a scholarship for academic excellence. Based on this information, rate the likelihood of the following outcomes from 1 (most likely) to 4 (least likely): A. Janet is a library science major.B. Janet turned 30 this year.C. Janet spends her weekends volunteering for the local museum.D. Janet is a library science major who spends her weekends volunteering for the local museum.	Conjunction Fallacy	D should be rated lower than A and C
6	A	Rep1f.04	Becky is going on a blind date with Robert. She doesn’t know much about him but she does know he is employed by the city; works variable, unpredicted hours; and loves adventure and seeks out adrenaline rushes. She also saw in his photo that he is quite burly. Based on this information, rate the likelihood of the following outcomes from 1 (most likely) to 4 (least likely): A. Robert is a city firefighter. B. Robert has traveled abroad.C. Robert likes action adventure movies.D. Robert is a city firefighter who likes action adventure movies.	Conjunction Fallacy	D should be rated lower than A and C
7	A	Rep2a.00	A group of 20 used car salespeople and 80 museum curators takes a personality questionnaire. You pick out one of the personality questionnaires and see that the person has personality traits that show he is extroverted and aggressive, places a high value on his appearance, and enjoys debates. Which of the following is more likely? A. He is a used car salesperson. B. He is a museum curator.	Base Rate Neglect	B
8	A	Rep2b.04	Consider a group in which 70% of the individuals are doctors and 30% are fashion models: an individual drawn at random from this group has been in seven different, highly publicized magazines. Which of the following is more likely: A. This person is a doctor. B. This person is a fashion model.	Base Rate Neglect	A
9	A	Rep2b.05		Base Rate Neglect	A

(continued on next page)

(*continued*)

Item	Subscale	Item Code	Item	Type	Answer
			Consider a group composed of 80% kangaroos and 20% tasmanian devils. An animal drawn at random has a very bad temperament, is easily startled, and has just bitten someone. Which of the following is more likely: A. This animal is a kangaroo B. This animal is a tasmanian devil		
10	A	Rep2b.09	Consider a group composed of 90 pilots and 10 carpenters. An individual is drawn at random who makes hand-crafted oak tables in his spare time. Which of the following is more likely: A. This individual is a pilot B. This individual is a carpenter	Base Rate Neglect	A
11	A	Rep2b.11	State University has 20,000 students. Approximately 4,000 of these students (20%) major in biology. A handful (20 out of 20,000 or 0.1%) major in Art History. Suppose we choose one student at random. Penelope is 20 years old and is described by her friends as somewhat impractical, emotional, and sensitive. She has traveled extensively in Europe and speaks French and Italian fluently. She is unsure about what career she will pursue on graduation, but she has demonstrated high levels of talent and won prizes for her calligraphy. On her boyfriend's last birthday, she wrote him a poem as a present. Which is more likely? A. Penelope is an Art History major? B. Penelope is a Biology major?	Base Rate Neglect	B
12	A	Rep2b.13	In a survey of staff at her middle school, Karen has counted 47 bus drivers and 3 dance teachers. She asked them all how they feel about dance performances, and one staff member answered with the following, "I love the vibrant energy I see in dancers, I attend performances regularly and I often take time to study new movement ideas in videos that people post to the internet. Which is more likely? A. This staff member is a bus driver. B. This staff member is a dance teacher.	Base Rate Neglect	A
13	A	Rep2c.02	Julian is a talented artist who has been praised by others for his creativity and ingenuity. 10 years from now, which of the following do you think is most likely? A. Julian works as a fashion designer for Gucci B. Julian works as an executive chef for a restaurant in Europe C. Julian is an assembly-line worker	Base Rate Neglect	C
14	A	Rep4a-2.26	Charlie is playing a dice game where players win points if the number rolled is higher that 3. The six-sided dice the players use are all fair, so there is a 50% chance than a given roll will be 3 or less, and a 50% chance the roll will be greater than 3. At one point, Charlie observes five dice rolls greater than 3 four times in a row. What do you think is the likelihood that the next dice roll will be less than 3? −100%, 75%, 50%, 25%, 0%	Misperception of Randomness	50%
15	A	Rep5.01	Suppose an unbiased coin is flipped 7 times, and each time the coin lands on heads. If you had to guess on the next toss, what side would you choose? A. Heads B. Tails C. Heads and tails are equally likely	Gambler's Fallacy	C
16	A	Rep5.06	The Hullaballoo Festival is held every year after the mayor of one of the two rival towns of Marysville and Burlingame draws one of two rabbits out of a hat. When a black rabbit is drawn, the festival is held in Marysville, and when a brown rabbit is drawn out of the hat, the festival is held in Burlingame. Over the past four years, a black rabbit has been drawn from the hat. Where is the festival more likely to be held this year? A. Burlingame. B. Marysville. C. Each town is equally likely to host the Hullaballoo Festival.	Gambler's Fallacy	C

(*continued on next page*)

(*continued*)

Item	Subscale	Item Code	Item	Type	Answer
17	A	Rep4a-2.01	During her morning commute, Molly drives through several intersections whose traffic lights are either red or green (there are no yellow lights in Molly's town). At any given moment, a light could be red or green with equal probability. One day, halfway through her commute, Molly runs into 5 red lights in a row. Which choice do you think most closely represents the likelihood that the next light Molly runs into is green? −100%, 75%, 50%, 25%, 0%	Misperception of Randomness	50%
18	A	Rep4a-2.09	The weather in Metropolis often varies but, overall, about half of the days are sunny (S) and half are rainy (R). For the past 7 days the weather has been rainy. What is the likelihood that tomorrow will be sunny? −100%, 75%, 50%, 25%, 0%	Misperception of Randomness	50%
19	A	Rep5.11	Phyllis is a good shot who can hit a target 100 yards away about 50% of the time. Today she is at the firing range and is on a roll–she's made her first 20 shots from 100 yards. What is the chance that she'll hit the target with her next shot? −100%, 75%, 50%, 25%, 0%	Misperception of Randomness	50%
20	A	Rep8-1.05	In her first semester at State University, Juliet takes 3 classes and really likes the instructor in each of her classes. Based on this, she assumes that the classes in each of her remaining 8 semesters at the university will also have wonderful instructors. How confident are you in Juliet's evaluation? 1 = Not confident at all; 7 = Very confident	Sample Size Neglect	The answer should be smaller than 4
21	A	Rep8-2.16	Micah's 10-year-old daughter Felicia scores two goals in her very first soccer game. Based on this, Micah proudly predicts that Felicia will be the top scorer for her team for the year (25 games). How confident are you in Micah's prediction? 1 = Not confident at all; 7 = Very confident	Sample Size Neglect	The answer should be smaller than 4
22	A	Rep8-2.12	Madeline and Denise love chocolate candy. Halloween comes and their mom puts together a candy bin. Whoever can guess how much chocolate is in the bin gets to keep it. Madeline tries first and scoops out a handful of candy–10 pieces. She notices that 1/2 are chocolate. Denise tries next and scoops out 30 pieces–1/3 are chocolate. Who is more likely to correctly guess how much chocolate candy is in the bin? A. Madeline B. Denise C. Both are equally likely	Sample Size Neglect	B
23	A	Rep8-1.01	Michael picks up a guitar and tries to play his favorite songs. He cannot hit a note and sounds awful. Because of this, Michael decides that he has no natural ability for music, and never tries to play an instrument again. How confident are you in Michael's evaluation? 1 = Not confident at all; 7 = Very confident	Sample Size Neglect	The answer should be smaller than 4
24	A	Rep8-1.02	Deborah orders furniture from IKEA for the first time. When it arrives, several pieces are broken, and a few are missing. Deborah concludes that IKEA furniture is shoddy and poor quality. How confident are you in her evaluation? 1 = Not confident at all; 7 = Very confident	Sample Size Neglect	The answer should be smaller than 4
25	A	Rep8-2.09	Melinda is a college volleyball coach interested in recruiting Mary for her team. For each item, rank how valuable the information should be in Melinda's decision of whether to recruit Mary on a scale from 1 (very high value) to 4 (very little value). A. Mary has been the best player the two times Melinda has watched her in person. B. Mary was the best player in the league during a recent 5-match play-off. C. Mary's parents were both elite volleyball players. D. Mary's average performance over 3 years (120 matches) places her in the top 15 in the state.	Sample Size Neglect	D should be ranked first
26	A	Rep8-2.05	At a carnival, Rick and Bobby come across a strange looking deck of cards. The description reads, "This deck contains either ¾ blue cards or ¼ blue cards. It is up to you to find out which!" Rick selects 6 cards from the deck and turns them over, revealing 5 blue cards and 1 green card. Bobby then snatches the deck away from Rick, shuffles the cards Rick pulled out back into the deck, and selects 20 cards at random, revealing 16 blue cards and 4 green cards. Both Rick and Bobby think that the deck contains ¾ blue cards. Who has better evidence to back up their claim? A. Rick has better evidence. B. Bobby has better evidence. C. They both have equally strong evidence	Sample Size Neglect	B

(*continued on next page*)

(*continued*)

Item	Subscale	Item Code	Item	Type	Answer
27	B	Rep2c.03	Jack's freshman year roommate provided a description of himself, including this excerpt, "I love to travel and investigate new places. People tell me I'm fun to talk to because I have a good memory for hilarious phrases and interesting stories. They also enjoy my music collection, especially the variety of non-English songs I've picked up over the years. Two years from now, which do you think is most likely? A. Jack's roommate chose to live in Japan. B. Jack's roommate became a comedian. C. Jack's roommate chose to study Business Administration.	Base Rate Neglect	C
28	B	Rep1d.00	Carrie is a very detail oriented college student. She likes putting together complex things, is technology-savvy, and is also very introverted. Please rank the follow in terms of probability (1 = highest probability, 4 = lowest probability): A. Carrie builds custom computers as a hobby.B. Carrie is a Psychology major.C. Carrie is a Psychology major who builds custom computers as a hobby.D. Carrie counsels high school students on career goals.	Conjunction Fallacy	C should be ranked lower than A and B
29	B	Rep1a.04	Betty is a passionate artist. She loves taking hikes outdoors and often takes long camping trips by herself to get away from the hectic stresses of her job. Which is more probable? A. Betty is a telemarketer. B. Betty is a telemarketer who leads a rock climbing group in her spare time.	Conjunction Fallacy	A
30	B	Rep1e.05	Sara is working on getting a Communications degree in college. She is very outgoing and loves being the center of attention. A. Sara is a cashier. B. Sara is a cashier and takes acting classes in her free time.	Conjunction Fallacy	A
31	B	Rep1e.08	Kenya has always loved fashion and spends most of her free time reading fashion magazines. Which is most likely? A. Kenya is a mother of two. B. Kenya is a mother of two who works at a fashion magazine.	Conjunction Fallacy	A
32	B	Rep2b.03	Consider a group of individuals in which 80% are recently released inmates and 20% are nurses: An individual drawn at random has turned in a lost wallet. Which of the following is more likely: A. This person is a former inmate B. This person is a nurse	Base Rate Neglect	A
33	B	Rep2b.06	Consider a group of individuals in which 90% are accountants and 10% are aeronautical engineers. An individual drawn at random spends most of her free time building models. A. This individual is an accountant B. This individual is an aeronautical engineer	Base Rate Neglect	A
34	B	Rep1e.00	Angela loves puzzles, works on mathematical proofs in her free time, and plans to get a graduate degree in math someday. Is it more likely that Angela is a waitress; is a waitress who also tutors college students in geometry? A. Angela is a waitress B. Angela is a waitress who also tutors college students in geometry	Conjunction Fallacy	A
35	**B**	Rep1f.02	Mark meets Tom at a mutual friend's house. Mark notices Tom appears to be of high intelligence, likes Star Trek, and makes lots of corny puns. Tom checks his email frequently on his phone. Based on this information, rate the likelihood of the following outcomes from 1 (most likely) to 4 (least likely): A. Tom is a software engineer.	Conjunction Fallacy	C should be ranked lower than A and B

(*continued on next page*)

(*continued*)

Item	Subscale	Item Code	Item	Type	Answer
			B. Tom spends his vacations at ComiCon.C. Tom is a software engineer who spends his vacations at ComiConD. Tom has the newest released cell phone available.		
36	B	Rep2a.02	A group of 100 children takes an intelligence evaluation. The group has 20 children who already know how to read and 80 children who have not yet learned to read. You pick out one of the evaluations and see that the child scored high on the intelligence evaluation. Which of the following is more likely? A. She already knows how to read. B. She has not yet learned how to read.	Base Rate Neglect	B
37	**B**	Rep2b.01	You are working on a yearbook for a local community college. 1000 students are graduating this year, 800 from the math program and 200 from the performing arts program. Each of the graduating students wrote a little blurb about themselves and you pick one at random. It reads: “My four years here flew by quickly! I met a lot of great people and made many new friends. I will truly miss getting my morning cup of coffee from the Starbucks on Thayer Street as well as snuggling up next to the fireplace in Morris Lounge with the latest Jack Kerouac novel. Professor Smith never ceases to amaze me with his witty banter and impeccable sense of humor.” Which of the following is more likely? A. The student is from the math program B. The student is from the liberal arts program	Base Rate Neglect	A
38	B	Rep2b.07	Consider a group composed of 90% Cats and 10% Dogs. An animal is drawn at random who tends to snarl and is very protective of his owner. Which of the following is more likely: A. This animal is a cat. B. This animal is a dog.	Base Rate Neglect	A
39	**B**	Rep2b.16	Miguel is at a company meeting of 200 people. Most (180) work in the factory but a few (20) are executives like him. He walks into a room and hears this sentence, “Our employees on the assembly line need to start working much harder if they want their bonus.” Which of the following is more likely? A. The individual speaking is an executive. B. The individual speaking works in the factory.	Base Rate Neglect	B
40	B	Rep4a-2.26	Charlie is playing a dice game where players win points if the number rolled is higher that 3. The six-sided dice the players use are all fair, so there is a 50% chance than a given roll will be 3 or less, and a 50% chance the roll will be greater than 3. At one point, Charlie observes five dice rolls greater than 3 four times in a row. What do you think is the likelihood that the next dice roll will be less than 3? −100%, 75%, 50%, 25%, 0%	Misperception of Randomness	50%
41	B	Rep5.01	Suppose an unbiased coin is flipped 7 times, and each time the coin lands on heads. If you had to guess on the next toss, what side would you choose? A. Heads B. Tails C. Heads and tails are equally likely	Gambler’s Fallacy	C
42	B	Rep5.06	The Hullaballoo Festival is held every year after the mayor of one of the two rival towns of Marysville and Burlingame draws one of two rabbits out of a hat. When a black rabbit is drawn, the festival is held in Marysville, and when a brown rabbit is drawn out of the hat, the festival is held in Burlingame. Over the past four years, a black rabbit has been drawn from the hat. Where is the festival more likely to be held this year? A. Burlingame. B. Marysville. C. Each town is equally likely to host the Hullaballoo Festival.	Gambler’s Fallacy	C
43	B	Rep4a-2.05	Cheryl has a deck of 26 playing cards. The deck has 13 spades and 13 clubs and the deck is properly shuffled. She draws a card from the deck and replaces that card in the deck, shuffling it again before selecting another card. After three draws from the deck Cheryl has drawn a club all three times in a row. Which choice do you think most closely represents the likelihood that the next card drawn from the deck will be a spade?	Misperception of Randomness	50%

(*continued on next page*)

(*continued*)

Item	Subscale	Item Code	Item	Type	Answer
			−100%, 75%, 50%, 25%, 0%		
44	B	Rep4a-2.08	Diane’s favorite game show tells participants that their final prize is behind one of three doors. Diane keeps track of which door the prizes were behind for each episode (1, 2, or 3), and many, many episodes later, has concluded that there is an equal probability (1/3) that the prize is behind a particular door. In a random selection of seven episodes, the prize is behind Door 2 five times in a row. What do you think is the likelihood that the prize will be behind Door 2 in the next episode? −100%, 66%, 33%, 10%, 0%	Misperception of Randomness	33%
45	B	Rep5.10	Randall is a good basketball player but generally only makes 25% of his free throws. However, in today’s game he has made his first 9 free throws. He’s now about to take his tenth free throw. What is the chance that he makes this shot? −100%, 75%, 50%, 25%, 0%	Misperception of Randomness	25%
46	B	Rep8-1.05	In her first semester at State University, Juliet takes 3 classes and really likes the instructor in each of her classes. Based on this, she assumes that the classes in each of her remaining 8 semesters at the university will also have wonderful instructors. How confident are you in Juliet’s evaluation? 1 = Not confident at all; 7 = Very confident	Sample Size Neglect	The answer should be smaller than 4
47	B	Rep8-2.16	Micah’s 10-year-old daughter Felicia scores two goals in her very first soccer game. Based on this, Micah proudly predicts that Felicia will be the top scorer for her team for the year (25 games). How confident are you in Micah’s prediction? 1 = Not confident at all; 7 = Very confident	Sample Size Neglect	The answer should be smaller than 4
48	B	Rep8-2.12	Madeline and Denise love chocolate candy. Halloween comes and their mom puts together a candy bin. Whoever can guess how much chocolate is in the bin gets to keep it. Madeline tries first and scoops out a handful of candy–10 pieces. She notices that 1/2 are chocolate. Denise tries next and scoops out 30 pieces–1/3 are chocolate. Who is more likely to correctly guess how much chocolate candy is in the bin? A. Madeline B. Denise C. Both are equally likely	Sample Size Neglect	B
49	B	Rep8-1.04	Peter brings home a dog (Rex) adopted from a local shelter. As soon as he walks in Rex spots a pair of slippers and immediately tears them to shreds. Peter now thinks he should return Rex because he is worried that the dog will always ruin his slippers. How confident are you in Peter’s evaluation? 1 = Not confident at all; 7 = Very confident	Sample Size Neglect	The answer should be smaller than 4
50	B	Rep8-2.07	Nate is interested in buying a car and consults several sources of information to determine how he should spend his money. Below are some of those sources of information. For each item, rank how valuable the information should be in his decision on a scale from 1 (very high value) to 4 (very little value). A. Nate’s neighbors each drive a Subaru and rave about its reliability and performance. They are both highly satisfied with their choice.B. Nate’s mechanic tells him that his business repairs fewer Subarus than most other makes of car and strongly recommends a Subaru.C. Many of Nate’s co-workers drive a Subaru. In individual discussions, they generally seem very happy with their car.D. In Consumer Reports’ survey of over 200,000 car-owners, Subarus get high ratings for reliability and performance.	Sample Size Neglect	D should be ranked first
51	B	Rep8-2.03	You are watching Shirley and Lucy practice at darts. Shirley throws 4 times and hits the bulls-eye twice. Lucy completes 100 throws and hits the bulls-eye 50 times. They are now getting ready for a match and you place a bet on who is more likely to hit the bulls-eye with half of their throws. Who should you bet on? A. Shirley B. Lucy C. They are equally likely to hit the bulls-eye	Sample Size Neglect	B
52	B	Rep8-2.11	Carl and Hank want to play dodgeball in the gym but to do so they need to get enough dodgeballs of the same color. Their gym teacher remembers that the ball bin contains either 2/7 yellow balls and 5/7 purple balls or 5/7 yellow balls and 2/7 purple balls but doesn’t remember which. To find out, Carl runs to the ball bin and pulls out 7 balls. Carl sees that 6 of the balls are purple and 1 of the balls is yellow, mixes the balls back in with the	Sample Size Neglect	B

(*continued on next page*)

(*continued*)

Item	Subscale	Item Code	Item	Type	Answer
			others, and runs to go tell his gym teacher. Hank runs to the ball bin after Carl is done and pulls out 30 balls. Hank sees that 20 of the balls are purple and 10 of the balls are yellow. Like Carl, Hank mixes the balls back in with the others and runs to go tell his gym teacher. Both Carl and Hank end up telling their gym teacher that the ball bin contains 2/7 yellow balls and 5/7 purple balls. Who has better evidence to back up their claim? A. Carl. B. Hank. C. They both have equally strong evidence		
53	C	Rep1a.02	Drew is 47, married, very outspoken, and intellectually sharp. He did not attend formal school past grade school. He is deeply concerned with getting the opportunity for higher education secured for his children. He has picketed for free education before he had children. A. Drew is a cab driver. B. Drew works on local political campaigns. C. Drew is a cab driver who works on local political campaigns.	Conjunction Fallacy	A or B
54	C	Rep1d.02	Lawrence is a fair-minded, disciplined individual who always waits for pedestrians to cross the street and often participates in volunteer opportunities with his children. Please rank the following in terms of probability (1 = highest probability, 5 = lowest probability): A. Lawrence is a local judgeB. Lawrence sells items at overly inflated prices on the internetC. Lawrence donates regularly to a local orphanageD. Lawrence is a used car salesmanE. Lawrence is a used car salesman who donates regularly to a local orphanage	ConjMultiR	E should be ranked lower than C and D
55	C	Rep1e.01	Liam feels energized when lots of things are going on, dreamt of being a firefighter when a child, and has a hard time sitting still. Which of the following is most likely? A. Liam is a plumber B. Liam is a plumber who volunteers with his local search and rescue team	Conjunction Fallacy	A
56	C	Rep1e.06	Bilbo has always been passionate about parties. He likes wearing silly costumes and having fun makeup on his face. A. Bilbo drives a Porsche. B. Bilbo drives a Porsche and is a stand-up comedian.	Conjunction Fallacy	A
57	C	Rep1e.07	Mike has few interests in life but his favorite thing to do is watch college football. Which is most likely? A. Mike drives a taxi. B. Mike drives a taxi and is a former high school football player.	ConjStd	A
58	C	Rep1f.00	You meet Sylvia at a friend's housewarming party. You learn in conversation that Sylvia likes to go hiking, camping, and fishing. You also find out used to be in the Peace Corps, and has lived in 8 different countries. Please rank the following in terms of probability (1 = highest probability, 4 = lowest probability) A. Sylvia works as a travel guide.B. Sylvia speaks three languages.C. Sylvia shops at sporting goods stores frequently.D. Sylvia works as a travel guide and shops at a sporting goods store frequently.	Conjunction Fallacy	C should be ranked lower than A and B
59	C	Rep2a.01	A group of 50 students participate in a research experiment on knowledge of statistics. Of the 50 students, 40 are from the Department of Anthropology, and 10 are from the Department of Finance. You choose a participant's answers and note that this student scored high in knowledge of statistics. This person is most likely to be: A. a Finance student. B. an Anthropology student.	Base Rate Neglect	B
60	C	Rep2a.03	A group of 75 student-athletes participate in a survey assessing knowledge of common sports injuries. Of the 75 student-athletes, 15 play football and 60 play baseball. You choose a participant's answers at random, and see the participant scored high in knowledge of head injuries. Which of the following is more likely?	Base Rate Neglect	A

(*continued on next page*)

(*continued*)

Item	Subscale	Item Code	Item	Type	Answer
			A. He plays football. B. He plays baseball.		
61	C	Rep2b.00	You are attending a conference with musicians and chefs from around the world. You know from reading the brochure that there are 90 chefs and 10 musicians giving talks. You pass by one of the conference rooms and hear someone talking about what it's like to perform on stage. Which of the following is more likely? A. The conference room's speaker is a chef. B. The conference room's speaker is a musician.	Base Rate Neglect	A
62	C	Rep2b.02	Consider a group in which 70% of the individuals are teachers for a living and 30% are firemen for a living: 10 people are in a room, 1 is randomly drawn. The man drawn has a very muscular build. He is a fearless individual who likes to take on new challenges. Which of the following is more likely: A. This person is a fireman. B. This person is a teacher.	Base Rate Neglect	A
63	C	Rep2b.10	Consider a group composed of 70 teachers and 30 lawyers. An individual is drawn at random who is opinionated, politically active, and aspires to be elected to the U.S. Senate. Which of the following is more likely: A. This individual is a teacher. B. This individual is a lawyer.	Base Rate Neglect	A
64	C	Rep2b.15	Jennifer is writing about volunteerism at her college. So far she has collected descriptions of volunteer experiences from 2 education students and 20 business students. You draw a description at random. "Helping middle schoolers with math homework has been a great volunteer experience for me and I really enjoy helping kids figure out how math works." Which of the following is more likely? A. This student is an education major. B. This student is a business major.	Base Rate Neglect	B
65	C	Rep2c.00	Tyrone is a short-tempered individual who has difficult getting along with others. He is currently in his junior year of high school. 10 years from now, which of the following do you think is most likely? A. Tyrone is an Olympic boxer B. Tyrone does data entry at a morgue C. Tyrone is a customer service representative for a major retail company	Base Rate Neglect	C
66	C	Rep4a-2.26	Charlie is playing a dice game where players win points if the number rolled is higher than 3. The six-sided dice the players use are all fair, so there is a 50% chance that a given roll will be 3 or less, and a 50% chance the roll will be>3. At one point, Charlie observes five dice rolls>3 four times in a row. What do you think is the likelihood that the next dice roll will be 3 or less? −100%, 75%, 50%, 25%, 0%	Misperception of Randomness	50%
67	C	Rep5.01	Suppose an unbiased coin is flipped 7 times, and each time the coin lands on heads. If you had to guess on the next toss, what side would you choose? A. Heads B. Tails C. Heads and tails are equally likely	Gambler's Fallacy	C
68	C	Rep5.06	The Hullaballoo Festival is held every year after the mayor of one of the two rival towns of Marysville and Burlingame draws one of two rabbits out of a hat. When a black rabbit is drawn, the festival is held in Marysville, and when a brown rabbit is drawn out of the hat, the festival is held in Burlingame. Over the past four years, a black rabbit has been drawn from the hat. Where is the festival more likely to be held this year?	Gambler's Fallacy	C

(*continued on next page*)

(*continued*)

Item	Subscale	Item Code	Item	Type	Answer
			A. Burlingame. B. Marysville. C. Each town is equally likely to host the Hullaballoo Festival.		
69	C	Rep4a-2.03	At a casino, Vanessa is watching her friends play roulette. The roulette wheel is a standard model where on each turn it can land on either a red or a black slot with equal probability. As Vanessa watches, she sees black come up 8 times in a row. Which choice do you think most closely represents the likelihood that the next slot the roulette lands on will be red?	Gambler's Fallacy	50%
			−100%, 75%, 50%, 25%, 0%		
70	C	Rep5.02	John plays for a minor league baseball team and normally gets a hit 1 out of every 4 times he bats (25% of the time). However, John has not had a hit in his last 10 at-bats. What is the chance that he will get a hit in his next at bat?	Gambler's Fallacy	25%
			−0%, 25%, 50%, 100%		
71	C	Rep5.03	Mark and Kim have 4 children–Mike, Matt, Luke, and Chris–all boys. Kim desperately wants a daughter and finally gets pregnant. She believes that her chances of having a girl are high considering she has had 4 boys in a row. How likely is it (as a percentage) that Kim will have a girl?	Gambler's Fallacy	50%
			−100%, 77%, 50%, 23%, 0%		
72	C	Rep8-1.05	In her first semester at State University, Juliet takes 3 classes and really likes the instructor in each of her classes. Based on this, she assumes that the classes in each of her remaining 8 semesters at the university will also have wonderful instructors. How confident are you in Juliet's evaluation?	Sample Size Neglect	The answer should be smaller than 4
			1 = Not confident at all; 7 = Very confident		
73	C	Rep8-2.16	Micah's 10-year-old daughter Felicia scores two goals in her very first soccer game. Based on this, Micah proudly predicts that Felicia will be the top scorer for her team for the year (25 games). How confident are you in Micah's prediction?	Sample Size Neglect	The answer should be smaller than 4
			1 = Not confident at all; 7 = Very confident		
74	C	Rep8-2.12	Madeline and Denise love chocolate candy. Halloween comes and their mom puts together a candy bin. Whoever can guess how much chocolate is in the bin gets to keep it. Madeline tries first and scoops out a handful of candy–10 pieces. She notices that 1/2 are chocolate. Denise tries next and scoops out 30 pieces–1/3 are chocolate. Who is more likely to correctly guess how much chocolate candy is in the bin?	Sample Size Neglect	B
			A. Madeline B. Denise C. Both are equally likely		
75	C	Rep8-1.03	When Martin was only six years old, his father bought him a pellet rifle. That same day, Martin shot and killed a rabbit. Martin's father speculates that he will grow up to be a very skilled hunter. How confident are you in his evaluation?	Sample Size Neglect	The answer should be smaller than 4
			1 = Not confident at all; 7 = Very confident		
76	C	Rep8-1.06	Marcela has been studying for years to be a surgeon. She was the top student in medical school and a top performer as a resident. She is hired by a hospital and, on the first day of the job, makes a careless error that nearly paralyzes a patient. Now she thinks she is not going to be a successful surgeon. How confident are you that Marcela's evaluation is accurate?	Sample Size Neglect	The answer should be smaller than 4
			1 = Not confident at all; 7 = Very confident		
77	C	Rep8-2.04	Monica has a meal at a local restaurant, Marvin's Steakhouse, but is disappointed that her steak was overcooked. Raul has been going to Marvin's for years and says that his steak has always been cooked just as he asked. You are now trying to decide whether to go to Marvin's for dinner. Who has better evidence on whether you should eat at Marvin's?	Sample Size Neglect	B
			A. Monica B. Raul C. They both have equally strong evidence		
78	C	Rep8-2.10		Sample Size Neglect	B

(*continued on next page*)

(*continued*)

Item	Subscale	Item Code	Item	Type	Answer
			Mark and Jacob are playing in a televised game show. There is a single bucket of balls in front of them. The game show host tells them that the bucket contains balls of two different colors, red and blue. 2/3 of the balls in the bucket are of one color and the remaining 1/3 of the balls are of the other color. The host then instructs Mark to pull 5 balls out of the bucket and examine their colors without showing Jacob. Mark sees that 4 of the balls are red and 1 of the balls is blue. The balls are then put back into the bucket and shuffled. Next, the host instructs Jacob to pull 20 balls out of the bucket and examine their colors without showing Mark. Jacob sees that 12 of the balls are red and 8 of the balls are blue. The balls are then put back in the bucket. The host then asks Mark and Jacob to guess whether the bucket contained 2/3 blue balls and 1/3 red balls or 2/3 red balls and 1/3 blue balls. Who has better evidence that the bucket contained 2/3 red and 1/3 blue? A. Mark B. Jacob C. They both have equally strong evidence		

References

Arkes, H. R. (1991). Costs and benefits of judgment errors: Implications for debiasing. *Psychological Bulletin, 110*(3), 486–498.

Bandura, A. (1977). *Social learning theory*. Englewood Cliffs, NJ: Prentice-Hall.

Barton, M., Symborski, C., Quinn, M. M., Morewedge, C. K., Kassam, K. S., & Korris, J. H. (2015). The use of theory in designing a serious game for the reduction of cognitive biases. *Paper presented at the DiGRA '15 - Proceedings of the 2015 DiGRA International Conference*.

Blandin, Y., & Proteau, L. (2000). On the cognitive basis of observational learning: Development of mechanisms for the detection and correction of errors. *Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology, 53*(3), 846–867. https://doi.org/10.1080/713755917.

Bollinger, B., & Gillingham, K. (2012). Peer effects in the diffusion of solar photovoltaic panels. *Marketing Science, 31*(6), 900–912. https://doi.org/10.1287/mksc.1120.0727.

Bong, C. L., Lee, S., Ng, A. S. B., Allen, J. C., Lim, E. H. L., & Vidyarthi, A. (2017). The effects of active (hot-seat) versus observer roles during simulation-based training on stress levels and non-technical performance: A randomized trial. *Advances in Simulation, 2*(1), 7–20. https://doi.org/10.1186/s41077-017-0040-7.

Brondani, M. A., & Rossoff, L. P. (2010). The "hot seat" experience: A multifaceted approach to the teaching of ethics in a dental curriculum. *Journal of Dental Education, 74*(11), 1220–1229.

Brown, I., & Inouye, D. (1978). Learned helplessness through modeling: The role of perceived similarity in competence. *Journal of Personality and Social Psychology, 36*(8), 900–908.

Chang, W., Chen, E., Mellers, B., & Tetlock, P. (2016). Developing expert political judgment: The impact of training and practice on judgmental accuracy in geopolitical forecasting tournaments. *Judgment and Decision Making, 11*(5), 509–526.

Choi, Y. J., & Luo, Y. (2015). 13-Month-olds' understanding of social interactions. *Psychological Science, 26*(3), 274–283. https://doi.org/10.1177/0956797614562452.

Cialdini, R. B., & Goldstein, N. J. (2004) Social influence: Compliance and conformity. Vol. 55. Annual Review of Psychology (pp. 591-621).

Cokely, E. T., Feltz, A., Ghazal, S., Allan, J. N., Petrova, D., & Garcia-Retamero, R. (2018). Skilled decision theory: From intelligence to numeracy and expertise. In K. A. Ericsson, R. R. Hoffman, A. Kozbelt, & A. M. Williams (Eds.), *The Cambridge handbook of expertise and expert performance* (2nd ed., pp. 476–505). New York, NY: Cambridge University Press.

Danchin, E., Giraldeau, L. A., Valone, T. J., & Wagner, R. H. (2004). Public information: From nosy neighbors to cultural evolution. *Science, 305*(5683), 487–491. https://doi.org/10.1126/science.1098254.

De Bruin, W. B., Parker, A. M., & Fischhoff, B. (2007). Individual differences in adult decision-making competence. *Journal of Personality and Social Psychology, 92*(5), 938–956. https://doi.org/10.1037/0022-3514.92.5.938.

Del Vicario, M., Bessi, A., Zollo, F., Petroni, F., Scala, A., Caldarelli, G., … Quattrociocchi, W. (2016). The spreading of misinformation online. *Proceedings of the National Academy of Sciences, 113*(3), 554–559. https://doi.org/10.1073/pnas.1517441113.

Del Vicario, M., Zollo, F., Caldarelli, G., Scala, A., & Quattrociocchi, W. (2017). Mapping social dynamics on Facebook: The Brexit debate. *Social Networks, 50*, 6–16. https://doi.org/10.1016/j.socnet.2017.02.002.

DellaVigna, S., & Kaplan, E. (2008). The political impact of media bias. In R. Islam (Ed.), Information and Public Choice: From Media Markets to Policymaking (pp. 79).

Epley, N., Morewedge, C. K., & Keysar, B. (2004). Perspective taking in children and adults: Equivalent egocentrism but differential correction. *Journal of Experimental Social Psychology, 40*(6), 760–768. https://doi.org/10.1016/j.jesp.2004.02.002.

Fischhoff, B. (1982). Debiasing. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 422–444). Cambridge, United Kingdom: Cambridge University Press.

Fleming, M. (1994). *Starting drama teaching*. London: David Fulton.

Fong, G. T., Krantz, D. H., & Nisbett, R. E. (1986). The effects of statistical training on thinking about everyday problems. *Cognitive Psychology, 18*(3), 253–292. https://doi.org/10.1016/0010-0285(86)90001-0.

Fong, G. T., & Nisbett, R. E. (1991). Immediate and delayed transfer of training effects in statistical reasoning. *Journal of Experimental Psychology: General, 120*(1), 34–45.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives, 19*(4), 25–42. https://doi.org/10.1257/089533005775196732.

Frith, C. D., & Frith, U. (2012). Mechanisms of social cognition. *Annual Review of Psychology, 63*(1), 287–313. https://doi.org/10.1146/annurev-psych-120710-100449.

Gergely, G., Bekkering, H., & Király, I. (2002). Rational imitation in preverbal infants. *Nature, 415*(6873), 755.

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology, 62*(1), 451–482. https://doi.org/10.1146/annurev-psych-120709-145346.

Gino, F. (2008). Do we listen to advice just because we paid for it? The impact of advice cost on its use. *Organizational Behavior and Human Decision Processes, 107*(2), 234–245. https://doi.org/10.1016/j.obhdp.2008.03.001.

Gino, F., & Moore, D. A. (2007). Effects of task difficulty on use of advice. *Journal of Behavioral Decision Making, 20*(1), 21–35. https://doi.org/10.1002/bdm.539.

Goldstein, N. J., Cialdini, R. B., & Griskevicius, V. (2008). A room with a viewpoint: Using social norms to motivate environmental conservation in hotels. *Journal of Consumer Research, 35*(3), 472–482. https://doi.org/10.1086/586910.

Hansen, M. (1999). The search-transfer problem: The role of weak ties in sharing knowledge across organization subunits. *Administrative Science Quarterly, 44*, 82–111.

Harvey, N., & Fischer, I. (1997). Taking advice: Accepting Help, improving judgment, and sharing responsibility. *Organizational Behavior and Human Decision Processes, 70* (2), 117–133. https://doi.org/10.1006/obhd.1997.2697.

Holmes, D. S. (1968). Dimensions of projection. *Psychological Bulletin, 69*(4), 248–268. https://doi.org/10.1037/h0025725.

Hoover, J. D., Giambatista, R. C., & Belkin, L. Y. (2012). Eyes on, hands on: Vicarious observational learning as an enhancement of direct experience. *Academy of Management Learning & Education, 11*(4), 591–608.

Huh, Y. E., Vosgerau, J., & Morewedge, C. K. (2014). Social defaults: Observed choices become choice defaults. *Journal of Consumer Research, 41*(3), 746–760.

Intelligence Advanced Research Projects Activity. (2011). Sirius Broad Agency Announcement, IARPA-BAA-11-03. Retrieved from http://www.iarpa.gov/index.php/research-programs/sirius/baa.

Intelligence Advanced Research Projects Activity. (2013). Unbiasing your biases II. Alexandria, VA: 522 Productions.

Jackson, V. A., & Back, A. L. (2011). Teaching communication skills using role-play: An experience-based guide for educators. *Journal of Palliative Medicine, 14*(6), 775–780. https://doi.org/10.1089/jpm.2010.0493.

John, O., Naumann, L., & Soto, C. (2008). Paradigm shift to the integrative Big Five trait taxonomy: History, measurement, and conceptual issues. In O. John, R. Robbins, & L. Pervin (Eds.), Handbook of personality: Theory and research (pp. 114–156). Guilford.

Kahneman, D. (2011). *Thinking, fast and slow*. New York, NY: Farrar, Straus and Giroux.

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology, 3*(3), 430–454. https://doi.org/10.1016/0010-0285(72)90016-3.

Kappes, H. B., & Morewedge, C. K. (2016). Mental simulation as substitute for experience. *Social and Personality Psychology Compass, 10*(7), 405–420. https://doi.org/10.1111/spc3.12257.

Kline, M. A. (2015). How to learn about teaching: An evolutionary framework for the study of teaching behavior in humans and other animals. *Behavioral and Brain Sciences, 38*. https://doi.org/10.1017/S0140525X14000090.

Kogut, B., & Zander, U. (1992). Knowledge of the firm, combinative capabilities, and the replication of technology. *Organization Science, 3*(3), 383–397.

Kraft-Todd, G. T., Bollinger, B., Gillingham, K., Lamp, S., & Rand, D. G. (2018). Credibility-enhancing displays promote the provision of non-normative public goods. *Nature, 563*(7730), 245–248. https://doi.org/10.1038/s41586-018-0647-4.

Krueger, J., & Stanke, D. (2001). The role of self-referent and other-referent knowledge in perceptions of group characteristics. *Personality and Social Psychology Bulletin, 27* (7), 878–888. https://doi.org/10.1177/0146167201277010.

Kuhlmeier, V., Wynn, K., & Bloom, P. (2003). Attribution of dispositional states by 12-month-olds. *Psychological Science, 14*(5), 402–408. https://doi.org/10.1111/1467-9280.01454.

Larrick, R. P., Morgan, J. N., & Nisbett, R. E. (1990). Teaching the use of cost-benefit reasoning in everyday life. *Psychological Science, 1*(6), 362–370. https://doi.org/10.1111/j.1467-9280.1990.tb00243.x.

Larrick, R. P., & Soll, J. B. (2006). Intuitions about combining opinions: Misappreciation of the averaging principle. *Management Science, 52*(1), 111–127. https://doi.org/10.1287/mnsc.1050.0459.

Lippman, S. A., & Rumelt, R. P. (1982). Uncertain imitability: An analysis of interfirm differences in efficiency under competition. *Bell Journal of Economics, 13*(2), 418–438.

Maddux, W. W., Yang, H., Falk, C., Adam, H., Adair, W., Endo, Y., … Heine, S. J. (). For whom is parting with possessions more painful? Cultural differences in the endowment effect. *Psychological Science, 21*(12), 1910–1917. https://doi.org/10.1177/0956797610388818.

Manz, C. C., & Sims, H. P. (1981). Vicarious learning: The influence of modeling on organizational behavior. *Academy of Management Review, 6*(1), 105–113. https://doi.org/10.2307/257144.

McEvily, S. K., & Chakravarthy, B. (2002). The persistence of knowledge-based advantage: An empirical test for product performance and technological knowledge. *Strategic Management Journal, 23*(4), 285–305. https://doi.org/10.1002/smj.223.

McNeil, B. J., Pauker, S. G., Sox, H. C. J., & Tversky, A. (1982). On the elicitation of preferences for alternative therapies. *New England Journal of Medicine, 306*(21), 1259–1262. https://doi.org/10.1056/NEJM198205273062103.

Mellers, B., Stone, E., Murray, T., Minster, A., Rohrbaugh, N., Bishop, M., … Tetlock, P. (2015). Identifying and cultivating superforecasters as a method of improving probabilistic predictions. *Perspectives on Psychological Science, 10*(3), 267–281. https://doi.org/10.1177/1745691615577794.

Milkman, K. L., Chugh, D., & Bazerman, M. H. (2009). How can decision making be improved? *Perspectives on Psychological Science, 4*(4), 379–383. https://doi.org/10.1111/j.1745-6924.2009.01142.x.

Miyamoto, Y., & Kitayama, S. (2002). Cultural variation in correspondence bias: The critical role of attitude diagnosticity of socially constrained behavior. *Journal of Personality and Social Psychology, 83*(5), 1239–1248. https://doi.org/10.1037//0022-3514.83.5.1239.

Mocanu, D., Rossi, L., Zhang, Q., Karsai, M., & Quattrociocchi, W. (2015). Collective attention in the age of (mis)information. *Computers in Human Behavior, 51*, 1198–1204. https://doi.org/10.1016/j.chb.2015.01.024.

Morewedge, C. K., Yoon, H., Scopelliti, I., Symborski, C. W., Korris, J. H., & Kassam, K. S. (2015). Debiasing decisions: Improved decision making with a single training intervention. *Policy Insights from the Behavioral and Brain Sciences, 2*(1), 129–140. https://doi.org/10.1177/2372732215600886.

Nadler, J., Thompson, L., & Boven, L. V. (2003). Learning negotiation skills: Four models of knowledge creation and transfer. *Management Science, 49*(4), 529–540.

Nelson, R. R., & Winter, S. G. (1982). *An evolutionary theory of economic change*. Boston: Belkman Press.

Nonaka, I., & von Krogh, G. (2009). Tacit knowledge and knowledge conversion: Controversy and advancement in organizational knowledge creation theory. *Organization Science, 20*(3), 635–652.

Nisbett, R. E. (1993). *Rules for reasoning*. Psychology Press.

Nisbett, R. E., Fong, G. T., Lehman, D. R., & Cheng, P. W. (1987). Teaching reasoning. *Science, 238*(4827), 625–631.

Northcraft, G. B., & Neale, M. A. (1987). Experts, amateurs, and real estate: An anchoring-and-adjustment perspective on property pricing decisions. *Organizational Behavior and Human Decision Processes, 39*(1), 84–97. https://doi.org/10.1016/0749-5978(87)90046-X.

Norton, M. I., Cooper, J., Monin, B., & Hogg, M. A. (2003). Vicarious dissonance: Attitude change from the inconsistency of others. *Journal of Personality and Social Psychology, 85*(1), 47–62. https://doi.org/10.1037/0022-3514.85.1.47.

Robbins, J. M., & Krueger, J. I. (2005). Social projection to ingroups and outgroups: A review and meta-analysis. *Personality and Social Psychology Review, 9*(1), 32–47. https://doi.org/10.1207/s15327957pspr0901_3.

Ross, L., Greene, D., & House, P. (1977). The "false consensus effect": An egocentric bias in social perception and attribution processes. *Journal of Experimental Social Psychology, 13*(3), 279–301.

Schkade, D. A., & Kahneman, D. (1998). Does living in California make people happy? A focusing illusion in judgments of life satisfaction. *Psychological Science, 9*(5), 340–346. https://doi.org/10.1111/1467-9280.00066.

Schultz, P. W., Nolan, J. M., Cialdini, R. B., Goldstein, N. J., & Griskevicius, V. (2007). The constructive, destructive, and reconstructive power of social norms: Research article. *Psychological Science, 18*(5), 429–434. https://doi.org/10.1111/j.1467-9280.2007.01917.x.

Schwitzgebel, E., & Cushman, F. (2015). Philosophers' biased judgments persist despite training, expertise and reflection. *Cognition, 141*, 127–137. https://doi.org/10.1016/j.cognition.2015.04.015.

Scopelliti, I., Min, H. L., McCormick, E., Kassam, K. S., & Morewedge, C. K. (2018). Individual differences in correspondence bias: Measurement, consequences, and correction of biased interpersonal attributions. *Management Science, 64*(4), 1879–1910. https://doi.org/10.1287/mnsc.2016.2668.

Scopelliti, I., Morewedge, C. K., McCormick, E., Min, H. L., Lebrecht, S., & Kassam, K. S. (2015). Bias blind spot: Structure, measurement, and consequences. *Management Science, 61*(10), 2468–2486. https://doi.org/10.1287/mnsc.2014.2096.

Sellier, A. L., Scopelliti, I., & Morewedge, C. K. (2019). Debiasing training transfers to improve decision making in the field. *Psychological Science, 30*(9), 1371–1379. https://doi.org/10.1177/0956797619861429.

Simmons, J. P., LeBoeuf, R. A., & Nelson, L. D. (2010). The effect of accuracy motivation on anchoring and adjustment: Do people adjust from provided anchors? *Journal of Personality and Social Psychology, 99*(6), 917–932. https://doi.org/10.1037/a0021540.

Soll, J. B., & Larrick, R. P. (2009). Strategies for revising judgment: How (and how well) people use others' opinions. *Journal of Experimental Psychology: Learning Memory and Cognition, 35*(3), 780–805. https://doi.org/10.1037/a0015145.

Soll, J. B., Milkman, K. L., & Payne, J. W. (2016). A User's Guide to Debiasing. In G. Keren, & G. Wu (Eds.), *Handbook of judgment and decision making*. Wiley-Blackwell.

Spencer, J. (2003). ABC of learning and teaching in medicine: Learning and teaching in the clinical environment. *British Medical Journal, 326*(7389), 591–594.

Strack, F., & Mussweiler, T. (1997). Explaining the enigmatic anchoring effect: Mechanisms of selective accessibility. *Journal of Personality and Social Psychology, 73* (3), 437–446. https://doi.org/10.1037/0022-3514.73.3.437.

Symborski, C., Barton, M., Quinn, M. M., Korris, J. H., Kassam, K. S., & Morewedge, C. K. (2017). The design and development of serious games using iterative evaluation. *Games and Culture, 12*(3), 252–268. https://doi.org/10.1177/1555412016673262.

Szulanski, G. (2000). The process of knowledge transfer: A diachronic analysis of stickiness. *Organizational Behavior and Human Decision Processes, 82*, 9–27.

Szulanski, G. (1996). Exploring internal stickiness: Impediments to the transfer of best practice within the firm. *Strategic Management Journal, 17*, 27–43.

Szulanski, Gabriel, Cappetta, Rossella, & Jensen, Robert J. (2004). When and how trustworthiness matters: Knowledge transfer and the moderating effect of causal ambiguity. *Organization Science, 15*(5), 600–613. https://doi.org/10.1287/orsc.1040.0096.

Tim, R. V. D., & Luthans, F. (1980). A social learning approach to organizational behavior. *The Academy of Management Review, 5*(2), 281–290. https://doi.org/10.2307/257438.

Tom, S. M., Fox, C. R., Trepel, C., & Poldrack, R. A. (2007). The neural basis of loss aversion in decision-making under risk. *Science, 315*(5811), 515–518. https://doi.org/10.1126/science.1134239.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. Biases in judgments reveal some heuristics of thinking under uncertainty. *Science, 185*(4157), 1124–1131.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science, 211*(4481), 453–458.

Von Hippel, E. (1994). "Sticky information" and the locus of problem solving: Implications for innovation. *Management Science, 40*(4), 429–439.

Wertz, A. E., & Wynn, K. (2014). Selective social learning of plant edibility in 6- and 18-month-old infants. *Psychological Science, 25*(4), 874–882. https://doi.org/10.1177/0956797613516145.

Wilson, T. D., Lindsey, S., & Schooler, T. Y. (2000). A model of dual attitudes. *Psychological Review, 107*(1), 101–126. https://doi.org/10.1037/0033-295X.107.1.101.
Yaniv, I. (2004). Receiving other people's advice: Influence and benefit. *Organizational Behavior and Human Decision Processes, 93*(1), 1–13. https://doi.org/10.1016/j.obhdp.2003.08.002.
Zander, U., & Kogut, B. (1995). Knowledge and the speed of the transfer and imitation of organizational capabilities: An empirical test. *Organization Science, 6*(1), 76–92.
Zimmerman, B. J., & Rosenthal, T. L. (1974). Observational learning of rule-governed behavior by children. *Psychological Bulletin, 81*(1), 29–42. https://doi.org/10.1037/h0035553.