RUNNING HEAD: Debiasing Decisions

Debiasing Decisions:

Improved Decision Making With A Single Training Intervention

Carey K. Morewedge and Haewon Yoon

Boston University, Questrom School of Business

Irene Scopelliti

City University London, Cass Business School

Carl W. Symborski

Leidos

James H. Korris

Creative Technologies Incorporated

Karim S. Kassam

Carnegie Mellon University, Dietrich School of Humanities and Social Sciences

Forthcoming: *Policy Insights from the Behavioral and Brain Sciences*

ABSTRACT

From failures of intelligence analysis to misguided beliefs about vaccinations, biased judgment and decision making contributes to problems in policy, business, medicine, law, and private life. Early attempts to reduce decision biases with training met with little success, leading scientists and policy makers to focus on debiasing by using incentives and changes in the presentation and elicitation of decisions. We report the results of two longitudinal experiments that found medium to large effects of one-shot debiasing training interventions. Participants received a single training intervention, played a computer game or watched an instructional video, which addressed biases critical to intelligence analysis (in Experiment 1: bias blind spot, confirmation bias, and fundamental attribution error; in Experiment 2: anchoring, representativeness, and social projection). Both kinds of interventions produced medium to large debiasing effects immediately (games $\geq$ -31.94% and videos $\geq$ -18.60%) that persisted at least 2 months later (games $\geq$ -23.57% and videos $\geq$ -19.20%). Games, which provided personalized feedback and practice, produced larger effects than did videos. Debiasing effects were domain-general: bias reduction occurred across problems in different contexts, and problem formats that were taught and not taught in the interventions. The results suggest that a single training intervention can improve decision making. We suggest its use alongside improved incentives, information presentation, and nudges to reduce costly errors associated with biased judgments and decisions.

Tweet: A single training intervention with an instructional game or video produced large and persistent reductions in decision bias.

Highlights:

- Biases in judgment and decision making create predictable errors in domains such as intelligence analysis, policy, law, medicine, business, and private life
- Debiasing interventions can be effective, inexpensive methods to improve decision making and reduce the costly errors that decision biases produce
- We found a short, single training intervention (i.e., playing a computer game or watching a video) produced persistent reductions in six cognitive biases critical to intelligence analysis
- Training appears to be an effective debiasing intervention to add to existing interventions such as improvements in incentives, information presentation, and how decisions are elicited (nudges)

> "Indeed, it appears that in some instances analysts' presumptions were so firm that they simply disregarded evidence that did not support their hypotheses. As we saw in several instances, when confronted with evidence that indicated Iraq did not have WMD, analysts tended to discount such information. Rather than weighing the evidence independently, analysts accepted information that fit the prevailing theory and rejected information that contradicted it. While analysts must adopt some frame of reference to interpret the flood of data they see, their baseline assumptions must be flexible enough to permit revision by discordant information. The analysts' frame of reference on Iraq's WMD programs—formed as it was by Iraq's previous use of such weapons, Iraq's continued efforts to conceal its activities, and Iraq's past success at hiding such programs—was so strong, however, that contradictory data was often discounted as likely false."
>
> - *Report to the President of the United States* (Silberman et al., 2005, p. 169)

Biased judgment and decision making is that which systematically deviates from the prescriptions of objective standards such as facts, rational behavior, statistics, or logic (Tversky & Kahneman, 1974). Decision bias is not unique to intelligence analysis. It affects the intuitions and calculated decisions of novices and highly trained experts in numerous domains including business, medicine, and law (Morewedge & Kahneman, 2010; Payne, Bettman, & Johnson, 1993) underlying phenomena such as the tendency to sell winning stocks too quickly and hold on to losing stocks too long (Shefrin & Statman, 1985), the persistent belief in falsified evidence linking vaccinations to autism (Lewandowsky, Ecker, Seifert, Schwarz, & Cook, 2012), and unintentional discrimination in hiring and promotion practices (Krieger & Fiske, 2006). Biased judgment and decision making affects people in their private lives. Less biased decision makers have more intact social environments, reduced risk of alcohol and drug use, lower childhood delinquency rates, and superior planning and problem solving abilities (Parker & Fischhoff, 2005).

Decision making ability varies across persons and within person across the lifespan (Bruine de Bruin, Parker, & Fischhoff, 2007; Dhami, Schlottmann, & Waldmann, 2011; Peters & Bruine de Bruin, 2011), but people are generally unaware of

the extent to which they are biased and have difficulty debiasing their decision making (Scopelliti, Morewedge, McCormick, Min, LeBrecht, & Kassam, 2015; Wilson & Brekke, 1994). Considerable scientific effort has been expended developing strategies and methods to improve novice and expert decision making over the last 50 years (for reviews, see Fischhoff, 1982; Soll Milkman, & Payne, in press). Three general debiasing approaches have been attempted, each with its pros and cons: changing incentives, optimizing choice architecture (e.g., improving how decisions are presented and elicited), and improving decision making ability through training.

INCENTIVES

Changing incentives can substantially improve decision making. Recalibrating incentives to reward healthy behavior improves diet (Schwartz, Mochon, Wyper, Maroba, Patel, & Ariely, 2014), exercise (Charness & Gneezy, 2009), weight loss (John et al., 2011), medication adherence (Volpp et al., 2008), and smoking cessation (Volpp et al., 2009). In one study, during a period in which the price of fresh fruit was reduced by 50% in suburban and urban school cafeterias, sales of fresh fruit increased four-fold (French, 2003). Incentives are not a solution for every bias, bias is prevalent even in high-stake multibillion-dollar decisions (Arkes & Blumer, 1985).

Incentives can also backfire. When incentives erode intrinsic motivation and change norms from prosociality to economic exchange, incentives demotivate behavior if they are insufficient or discontinued (Gneezy Meier, & Rey-Biel, 2011). Israeli daycare facilities that introduced a small fine when parents picked up their children late, for instance, saw an *increase* in the frequency of late pickups. The fine made rude behavior

acceptable, a price to watch the children a little longer (Gneezy & Rustichini, 2000). When incentives are too great, they can make people choke under pressure (Ariely, Gneezy, Loewenstein, & Mazar, 2009). If people apply inappropriate decision strategies or correction methods because they do not know how or the extent to which they are biased, increasing incentives can exacerbate bias rather than mitigate it (Lerner & Tetlock, 1999). In short, incentives can effectively improve behavior, but they require careful calibration and implementation.

OPTIMIZING CHOICE ARCHITECTURE

Optimizing the structure of decisions, how choice options are presented and how choices are elicited, is a second way to effectively debias decisions. People do make better decisions when they have the information they need and good options to choose from. Giving people more information and choices is not always helpful, particularly when it makes decisions too complex to comprehend, existing biases encourage good behavior, or people recognize the choices they need to make but fail to implement them because they lack self-control (Bhargava & Loewenstein, 2015; Fox & Sitkin, 2015). Providing calorie information does not necessarily lead people to make healthier food choices, for instance, and there is some evidence that smokers actually overestimate the health risks of smoking—debiasing smokers may actually increase their health risks (Downs, Loewenstein, & Wisdom, 2009).

Changing what and how information is presented can make choices easier to understand and good options easier to identify, thus doing more to improve decisions than simply providing more information. Eligible taxpayers are more likely to claim their

Earned Income Tax Credits, for example, when benefit information is simplified and prominently displayed (e.g., "…of up to $5,657"; Bhargava & Manoli, in press). Consumers are better able to recognize that small reductions in the fuel consumption of inefficient vehicles saves more fuel than large reductions in the fuel consumption of efficient vehicles (e.g., improving 16MPG to 20MPG saves more than improving 34MPG to 50MPG) when the same information about vehicle fuel consumption is framed in gallons per 100 miles (GPM) rather than in MPG (Larrick & Soll, 2008). Both novices and trained experts benefit from the implementation of simple visual representations of risk information, whether they are evaluating medical treatments or new counterterrorism techniques (Garcia-Retamero & Dhami, 2011; 2013). Moreover, statistical analyses of voting patterns in the 2000 United States Presidential Election suggest that had the butterfly ballots used by Palm Beach County, Florida been designed in a manner not inconsistent with basic principles of perception, Al Gore would have been elected President (Fox & Sitkin, 2015).

Even when people fully understand their options, if one option is better for them or society but choosing it requires effort, expertise, or self-control, its selection can be increased if small nudges in presentation and elicitation methods are implemented (Thaler & Sunstein, 2008). Nudges take many forms such as information framing, commitment devices, and default selection. Voters are more mobilized by message frames that emphasize a high expected turnout at the polls (implying voting is normative) than message frames that emphasize low expected turnouts (implying each vote is important; Gerber & Rogers, 2009), and consumers prefer lower-fat meat when its fat content is framed as 25% fat than 75% lean (Levin & Gaeth, 1988). Shoppers are willing

to commit to foregoing cash rebates that they currently receive on healthy foods if they fail to increase the amount of healthy food that they purchase by 5% (Schwartz et al., 2014), and employees substantially increase their contributions to 401k programs when they commit to allocating money from future raises to their retirement savings before receiving those raises (Thaler & Benartzi, 2004).

People are more likely to choose an option if it is a default from which they must opt-out than if it is an option that they must actively choose (i.e., "opt-in"). In one study, university employees were 36% more likely to receive a flu shot if emailed an appointment from which they could opt-out, than if emailed a link from which they could schedule an appointment (Chapman, Li, Colby, & Yoon, 2010). Organ donation rates are at least 58% higher in European countries in which the default is to opt-out of being a donor than in which the default is to opt-in (Johnson & Goldstein, 2003). Selecting better default options is not necessarily coercive. It results in outcomes that decision makers themselves prefer (Goldstein, Johnson, Herrman, & Heitmann, 2008; Huh, Vosgerau, & Morewedge, 2014).

The potential applications of optimizing of choice architecture are broad, ranging from increasing retirement savings and preserving privacy, to reducing the gasoline, soda, and junk food that people consume (Acquisti, Brandimarte, & Loewenstein, 2015; Larrick & Soll, 2008; Schwartz et al., 2014; Thaler & Benartzi, 2004). Optimizing choice architecture is a cheap way to improve public welfare while preserving freedom of choice, as it does not exclude options or change economic incentives (Camerer, Issacharoff, Loewenstein, O'Donoghue, & Rabin, 2003; Thaler & Sunstein, 2003; 2008). Critics, however, point out that these improvements may not do enough. They tend to

reduce decision bias in one, not multiple contexts, and do not address the underlying structural causes of biased decisions such as poorly calibrated incentives or bad options (Bhargava & Loewenstein, 2015).

TRAINING

Training interventions to improve decision making, to date, have met with limited success mostly in specific domains. Training can be very effective when accuracy requires experts to recognize patterns and select an appropriate response, such as in weather forecasting, firefighting, and chess (Phillips, Klein, & Siek, 2004). By contrast, even highly trained professionals are less accurate than very simple mathematical models in other domains such as parole decisions, personnel evaluations, and clinical psychological testing (Dawes, Faust, & Meehl, 1989). Whether domain-specific expertise is achievable appears to be contingent on external factors such as the prevalence of clear feedback, the frequency of the outcome being judged, and the number and nature of variables that determine that outcome (Kohler, Brenner, & Griffin, 2002; Harvey, 2011).

Evidence that training effectively improves general decision making ability is inconclusive at present (Arkes, 1991; Milkman, Chugh, & Bazerman, 2009; Phillips et al., 2004). Weather forecasters are well calibrated when predicting the chance of precipitation (Murphy & Winkler, 1974), for example, but are overconfident in their answers to general knowledge questions (Wagenaar & Keren, 1986). Even within their domain of expertise, experts struggle to apply their training to new problems. Philosophers trained in logic exhibit the same preference reversals in similar moral dilemmas as academics without logic training (Schwitzgebel & Cushman, 2012), and

physicians exhibit the same preference reversals as untrained patients for equivalent medical treatments when those treatments are framed in terms of survival or mortality rates (McNeil, Paulker, Sox, & Tversky, 1982). Several studies have shown that people do not apply their training to unfamiliar and dissimilar domains because they lack the necessary metacognitive strategies to recognize underlying problem structure (for reviews, see Barnett & Ceci, 2002; Reeves & Weisberg, 1994; Willingham, 2008).

Debiasing training methods teaching inferential rules (e.g., "consider-the-opposite" and "consider-an-alternative" strategies) that are grounded in two-system models of reasoning hold some promise (e.g., Lilenfield et al., 2009; Milkman, Chugh, & Bazerman, 2009; Soll, Milkman, & Payne, in press). Two-system models of reasoning assume that people initially make an automatic intuitive judgment that can be subsequently accepted, corrected, or replaced by more controlled and effortful thinking: through "System 1" and "System 2" processes, respectively (Evans, 2003; Morewedge & Kahneman, 2010; Sloman, 1996). Recognizing that "1593 x 1777" is a math problem and that its answer is a large number, for instance, are automatic outputs of System 1 processes. Deducing the answer to the problem requires the engagement of effortful System 2 processes.

Effective debiasing training typically encourages the consideration of information that is likely to be underweighted in intuitive judgment (e.g., Hirt & Markman, 1995), or teaches people statistical reasoning and normative rules of which they may be unaware (e.g., Larrick, Morgan, & Nisbett, 1990). In large doses, debiasing training can be effective. Coursework in statistical reasoning, and graduate training in probabilistic sciences such as psychology and medicine, does appear to increase the use of statistics

and logic when reasoning about everyday problems to which they apply (Nisbett et al., 1987).

PERSISTENT DEBIASING WITH A SINGLE INTERVENTION

We tested whether a single debiasing training intervention could effectively produce immediate and persistent improvements in decision making. In two experiments, we directly compared the efficacy of two debiasing training interventions, a video and an interactive serious (i.e., educational) computer game. Videos and games are scalable training methods that can be used for efficient teaching of cognitive skills (e.g., Downs, 2014; Haferkamp, Kraemer, Linehan, & Schembri, 2011; Sliney & Murphy, 2008). The experiments, funded by Intelligence Advanced Research Projects Activity BAA-11-03, tested whether debiasing training could produce persistent reductions in six cognitive biases identified by our program sponsor as affecting all types of intelligence analysis.

Experiment 1 targeted three cognitive biases: *bias blind spot* (i.e., perceiving oneself to be less biased than one's peers; Scopelliti et al., 2015), *confirmation bias* (i.e., gathering and interpreting evidence in a manner confirming rather than disconfirming the hypothesis being tested; Nickerson, 1998), and *fundamental attribution error* (i.e., attributing the behavior of a person to dispositional rather than to situational influences; Gilbert, 1998; Jones & Harris, 1967). Experiment 2 targeted three different cognitive biases: *anchoring* (i.e., overweighting the first information primed or considered in subsequent judgment; Tversky & Kahneman, 1974), bias induced by over-reliance on *representativeness* (i.e., using the similarity of an outcome to a prototypical outcome to judge its probability; Kahneman & Tversky, 1972), and social *projection* (i.e., assuming

others' emotions, thoughts, and values are similar to one's own; Epley, Morewedge, & Keysar, 2004; Robbins & Krueger, 2005).

Many tasks crucial to intelligence analysis are influenced by these biases (for a review, see Heuer, 1999). Analysts must assess evidence with uncertain truth value (e.g., anchoring, bias blind spot, confirmation bias). They must infer cause and effect when evaluating past, present, and future events (e.g., confirmation bias, representativeness), the behavior of persons, and the actions of nations (e.g., fundamental attribution error, projection). Analysts regularly estimate probabilities (e.g., anchoring, confirmation bias, projection bias, representativeness), evaluate their own analyses, and evaluate the analyses of others (e.g., anchoring, bias blind spot, confirmation bias, projection bias). Although each of these cognitive biases may have its unique influence, multiple biases are likely to act in concert in any complex assessment (Cooper, 2005).

Attempting to reduce these biases with videos and games allowed us to administer short, one-shot training interventions (i.e., approximately 30 and 60 minutes, respectively) using two different mixes of the four debiasing training procedures proposed by Fischhoff (1982): (1) teaching people about each bias, (2) teaching people the directional influence of each bias on judgment, (3) providing feedback, and (4) providing extended feedback with coaching, intervention, and mitigating strategies. The videos incorporated debiasing training procedures 1, 2, and mitigating strategies (i.e., 4 without feedback, intervention, or coaching) in a passive format. The games incorporated all four debiasing training procedures in an interactive format. Each participant watched one video or played one game, without repetition.

Each video instructed viewers about three cognitive biases, gave examples of each bias, and provided mitigating strategies (e.g., consider alternative explanations, anchors, possible outcomes, perspectives, base-rates, countervailing evidence, and potential situational influences on behavior). Each of the interactive computer games elicited the same three cognitive biases during gameplay by asking players to make in-game decisions based on limited evidence (e.g., testing a hypothesis, evaluating the behavior of a character in the game, etc.). In an after-action review (AAR) at the end of each of three levels of each game, players were given definitions and examples of the three biases, personalized feedback on the degree to which they exhibited each bias, and mitigating strategies and practice. Like the video, the mitigating strategies taught in the game included: consider alternative explanations, anchors, possible outcomes, perspectives, base-rates, countervailing evidence, and consider potential situational influences on behavior. In addition, the games taught formal rules of logic (e.g., the conjunction of two events can be no more likely than either event on its own), methods of hypothesis testing (e.g., hold all variables other than the suspected causal variable constant when testing a hypothesis), and relevant statistical rules (e.g., large samples are more accurate representations than small samples), as well as encouraging participants to carefully reconsider their initial answers.

Our experiments tested the immediate and persistent effects of the debiasing interventions by measuring the extent to which participants committed each bias three times: in a pretest before training, in a posttest immediately after training, and in follow-up testing 8 or 12 weeks after training (see Figure 1). The pretest, training, and posttest were conducted in our laboratory and measured immediate debiasing effects of the

training interventions. The follow-up was administered online and measured the persistent debiasing effects of the training interventions over a longer term. Sample sizes were declared in advance to our government sponsor, and independent third-party analyses of the data were performed that confirmed the accuracy of our results (Kopecky, McKneely, & Bos, 2015).

EXPERIMENT 1: BIAS BLIND SPOT, CONFIRMATION BIAS, AND FUNDAMENTAL ATTRIBUTION ERROR

Method

Participants

Two hundred and seventy-eight people in a convenience sample recruited in Pittsburgh, PA (132 women; M_{age} = 24.5, *SD* = 8.52) received $30 for completing a laboratory training session, and an additional $30 payment for completing a follow-up test online. Most (80.2%) participants had some college education, 14.3% had graduate or professional degrees. A total of 243 participants successfully completed the laboratory portion of the experiment (Game *n* = 160; Video *n* = 83); 196 successfully completed the online follow-up (Game *n* = 130; Video *n* = 66).[2]

Training Interventions

Video. Unbiasing Your Biases is a 30-minute unclassified training video (produced by Intelligence Advanced Research Projects Activity, 2012). A narrator first defines heuristics and explains how heuristics can sometimes lead to incorrect inferences. He then defines bias blind spot, confirmation bias, and fundamental attribution error,

presents vignettes in which actors commit each bias, gives an additional example of fundamental attribution error and confirmation bias, and suggests mitigating strategies. The last two minutes of the video is a comprehensive review of its content.

Game. Missing: The Pursuit of Terry Hughes is a computer game designed to elicit and mitigate bias blind spot, confirmation bias, and fundamental attribution error (produced by Symborski, Barton, Quinn, Morewedge, Kassam, & Korris, 2014). It is a first person point-of-view educational game, in which the player searches for a missing neighbor (i.e., Terry Hughes) and exonerates her of criminal activity. During interactive gameplay in each of three levels, players make judgments designed to test the degree to which they exhibit confirmation bias and the fundamental attribution error. After-action reviews at the end of each level feature experts explaining each bias and narrative examples. To elicit bias blind spot, players then assess their degree of bias during each level. Next, participants are given personalized feedback on the degree of bias they exhibited. Finally, participants perform additional practice judgments of confirmation bias (5 in total) and receive immediate feedback before the next level begins or the game ends.[1]

Bias Measures

We developed measures of the extent to which participants committed each of the three cognitive biases: a Bias Blind Spot scale (BBS), a Fundamental Attribution Error scale (FAE), and six Confirmation Bias scales (CB). These were tested to ensure reliability and validity (see Supplemental Materials). Three interchangeable version of each scale (i.e., subscales) were created to measure bias commission at pretest, posttest,

and follow-up. Scoring of each subscale ranged from 0 (no biased answers) to 100 (all answers biased). Confirmation bias scale scores were calculated by averaging the six CB scales at pretest, posttest, and follow-up. Overall bias commission scores at pretest, posttest, and follow-up were calculated by averaging the three bias subscale scores at that time point (i.e., BBS, FAE, CB).

Ancillary scales measuring bias knowledge were developed to assess changes in ability to recognize instances of the three biases and discriminate between them. Bias knowledge scales were scored on a 0-100 scale, with higher scores indicating greater ability to recognize and discriminate between the three biases.

Testing Procedure

In a laboratory session, each participant was seated in a private cubicle with a computer. Participants first completed the pretest measure, consisting of three subscales assessing their commission of each of the three cognitive biases (i.e., BBS, CB, and FAE). Participants also completed a bias knowledge scale at this time. Next, each participant was randomly assigned to receive one of the training interventions, to either play the game or watch the video, without repetition. Immediately after training, participants completed the posttest measure, consisting of three subscales assessing their commission of each of the three cognitive biases post-training (i.e., BBS, CB, and FAE). Participants also completed a bias knowledge posttest at this time. To measure the persistence of debiasing training, eight weeks from the day in which he or she completed the laboratory session, each participant received a personalized link via email to complete the follow-up measure, consisting of three subscales assessing his or her commission of

each of the three biases (i.e., BBS, CB, and FAE). He or she had seven days to complete the follow-up measure in one sitting. Participants also completed a bias knowledge measure at this time. The specific bias scales serving as the pretest, posttest, and follow-up measures of bias commission and bias knowledge were counterbalanced across participants.

Results

Scale Reliability

Subscales were reliable. Bias blind spot (Cronbach's α): $.77_{pretest}$, $.82_{posttest}$, and $.76_{follow\text{-}up}$. Confirmation bias: $.73_{pretest}$, $.73_{posttest}$, and $.76_{follow\text{-}up}$. FAE: $.68_{pretest}$, $.77_{posttest}$, and $.78_{follow\text{-}up}$.

Bias Commission

Main effects of training on bias commission overall and for each of the three cognitive biases were analyzed using 2 (training: game vs. video) x 2 (timing: pretest vs. posttest or pretest vs. follow-up) mixed ANOVAs with repeated measures on the last factor. To compare the efficacy of the game and video, between subjects (training: game vs. video) ANCOVAs were performed to compare the debiasing effects of the training methods at posttest and follow-up, controlling for pretest scores. Means of bias commission scores for overall bias and each of the three biases by training intervention conditions are presented in Figure 2 (bias knowledge scores are only reported in the text).

Overall bias. Overall, training effectively reduced cognitive bias immediately and two months later, $F(1, 241) = 439.23$, $p < .001$ and $F(1, 194) = 179.88$, $p < .001$, respectively. Debiasing effect sizes (Rosenthal & Rosnow, 1991) for overall bias were

large for the game ($d_{\text{pre-post}} = 1.68$ and $d_{\text{pre-followup}} = 1.11$) and medium for the video ($d_{\text{pre-post}} = .69$ and $d_{\text{pre-followup}} = .66$). The game more effectively debiased participants than did the video immediately and two months later, $F(1, 240) = 68.8$, $p < .001$ and $F(1, 193) = 12.69$, $p < .001$, respectively.

Bias blind spot. Training effectively reduced BBS immediately and two months later, $F(1, 241) = 151.66$, $p < .001$ and $F(1, 194) = 104.51$, $p < .001$, respectively. Debiasing effect sizes for BBS were large for the game ($d_{\text{pre-post}} = .98$ and $d_{\text{pre-followup}} = .89$) and medium for the video ($d_{\text{pre-post}} = .49$ and $d_{\text{pre-followup}} = .49$). The game more effectively debiased participants than did the video immediately and two months later, $F(1, 240) = 17.31$, $p < .001$ and $F(1, 193) = 13.18$, $p < .001$, respectively.

Fundamental attribution error. Training effectively reduced FAE immediately and two months later, $F(1, 241) = 183.74$, $p < .001$ and $F(1, 194) = 85.32$, $p < .001$, respectively. Debiasing effect sizes for FAE were large and medium for the game ($d_{\text{pre-post}} = 1.12$ and $d_{\text{pre-followup}} = .72$) and medium and small for the video ($d_{\text{pre-post}} = .38$ and $d_{\text{pre-followup}} = .52$). The game more effectively debiased participants than did the video immediately and two months later, $F(1, 240) = 50.06$, $p < .001$ and $F(1, 193) = 6.53$, $p < .05$, respectively.

Confirmation bias. Training effectively reduced confirmation bias immediately and two months later, $F(1, 241) = 181.08$, $p < .001$ and $F(1, 194) = 45.52$, $p < .001$, respectively. Debiasing effect sizes for confirmation bias were large to medium the game ($d_{\text{pre-post}} = 1.09$ and $d_{\text{pre-followup}} = .58$) and medium to small for the video ($d_{\text{pre-post}} = .38$ and $d_{\text{pre-followup}} = .26$). The game more effectively debiased participants than did the video

immediately and two months later, $F(1, 240) = 33.54$, $p < .001$ and $F(1, 193) = 5.17$, $p < .05$, respectively.

Our scales tested six different facets of confirmation bias, but our game only taught three. This testing structure allowed us to test the generalization of debiasing training across trained (Snyder & Swann, 1978; Tschirgi, 1980; Wason, 1960) and untrained facets of confirmation bias (Downs & Shafir, 1999; Nisbett & Ross, 1980; Wason, 1968). Compared to their pretest scores, participants exhibited a reduction in confirmation bias on the trained facets at posttest and follow-up, $t(159) = 9.81$, $p < .001$, $d = .78$ and $t(129) = 2.69$, $p < .01$, $d = .24$, respectively. More important, compared to their pretest scores, participants exhibited reduced confirmation bias for untrained facets at posttest and follow-up, $t(159) = 10.05$, $p < .001$, $d = .79$ and $t(129) = 7.42$, $p < .001$, $d = .65$, respectively. Controlling for their pretest scores, participants performed better on trained than untrained facets of confirmation bias at posttest, $t(159) = 2.56$, $p < .05$, $d = .20$, but there were no significant differences between trained and untrained facets at follow-up, $t < 1$ (for means, see Figure 3).

Bias Knowledge

Training also effectively improved bias knowledge immediately and two months later, $F(1, 241) = 385.13$, $p < .001$ and $F(1, 194) = 64.31$, $p < .001$, respectively. Bias knowledge increased for participants who played the game ($M_{pretest} = 35.78$, $M_{posttest} = 58.54$, $M_{follow\text{-}up} = 47.98$, $d_{pre\text{-}post} = 1.05$ and $d_{pre\text{-}followup} = .52$) and watched the video ($M_{pretest} = 35.29$, $M_{posttest} = 69.28$, $M_{followup} = 50.63$, $d_{pre\text{-}post} = 1.69$ and $d_{pre\text{-}follow\text{-}up} = .69$). The video more effectively taught participants to recognize and discriminate bias than did

the game immediately, $F(1, 240) = 15.52$, $p < .001$, but was no more effective two months later, $F < 1$.

EXPERIMENT 2: ANCHORING, PROJECTION BIAS, AND REPRESENTATIVENESS

Method

Participants

Two hundred and sixty-nine people in a convenience sample recruited in Pittsburgh, PA (155 women; $M_{age} = 27.8$, $SD = 12.01$) received $30 for completing a laboratory training session, and an additional $30 payment for completing a follow-up test online. Most (94.1%) participants had some college education, 19.3% had graduate or professional degrees. A total of 238 participants successfully completed the laboratory portion of the experiment (Game $n = 156$; Video $n = 82$); 192 successfully completed the online follow-up (Game $n = 126$; Video $n = 66$).[2]

Stimuli

Training video. Unbiasing Your Biases 2 (Intelligence Advanced Research Projects Activity, 2013) had the same structure as the video in Experiment 1, but addressed anchoring, projection, and representativeness.

Computer Game. Missing: The Final Secret is a serious game designed to elicit and mitigate to anchoring, projection, and representativeness. The game followed a narrative arc, genre, and structure similar to the game in Experiment 1 (see Barton, Symborski, Quinn, Morewedge, Kassam, & Korris, 2015). Players exonerate their employer of a criminal charge and uncover the criminal activity of her accusers, while

making decisions testing their commission of each of the cognitive biases during game play. Experiment 2 introduced adaptive training in the AARs. When players gave biased answers to practice questions, they received additional practice questions (up to 16 in total) and feedback.[1]

Scale Development

Scales measuring commission of anchoring, projection, and representativeness, and scales measuring bias knowledge were developed and scored following a procedure similar to that used in Experiment 1 (see Supplemental Materials).

Testing Procedure

The experiment adhered to the same testing procedure as described in Experiment 1, with the exception that the follow-up was administered 12 weeks after participants completed their laboratory session.

Results

Scale Reliability

Subscales were reliable: Anchoring (Cronbach's α): $.60_{pretest}$, $.52_{posttest}$, and $.62_{follow\text{-}up}$. Projection bias: $.63_{pretest}$, $.78_{posttest}$, and $.77_{follow\text{-}up}$. Representativeness: $.86_{pretest}$, $.87_{posttest}$, and $.93_{follow\text{-}up}$.

Bias Commission

The same analyses were performed as in Experiment 1. All bias commission scale means are presented in Figure 2.

Overall Bias. Overall, training effectively reduced cognitive bias immediately and three months later, $F(1, 236) = 719.58$, $p < .001$ and $F(1, 190) = 246.17$, $p < .001$, respectively. Debiasing effect sizes for overall bias were large for both the game ($d_{\text{pre-post}} = 1.74$ and $d_{\text{pre-followup}} = 1.16$) and video ($d_{\text{pre-post}} = 1.75$ and $d_{\text{pre-followup}} = 1.07$). However, the game more effectively debiased participants than did the video immediately, $F(1, 235) = 13.44$, $p < .001$, and marginally three months later, $F(1, 189) = 3.66$, $p = .057$.

Anchoring. Training effectively reduced anchoring immediately and three months later, $F(1, 236) = 127.94$, $p < .001$ and $F(1, 190) = 78.42$, $p<.001$, respectively. Debiasing effect sizes for anchoring were medium for the game ($d_{\text{pre-post}} = .70$ and $d_{\text{pre-followup}} = .63$) and large to medium for the video ($d_{\text{pre-post}} = .80$ and $d_{\text{pre-followup}} = .66$). The game and video were equally effective immediately and three months later, Fs < 1, ps > .62.

Projection. Training effectively reduced projection immediately and three months later, $F(1, 236) = 197.29$, $p < .001$ and $F(1, 190) = 34.52$, $p < .001$, respectively. Debiasing effect sizes for projection were large to medium for the game ($d_{\text{pre-post}} = 1.11$ and $d_{\text{pre-followup}} = .54$) and medium to small for the video ($d_{\text{pre-post}} = .49$ and $d_{\text{pre-followup}} = .14$). The game more effectively debiased participants than did the video immediately and three months later, $F(1, 235) = 34.42$, $p < .001$ and $F(1, 189) = 13.49$, $p < .001$, respectively.

Representativeness. Training effectively reduced bias due to overreliance on representativeness immediately and three months later, $F(1, 236) = 599.55$, $p < .001$ and $F(1, 190) = 216.36$, $p < .001$, respectively. Debiasing effect sizes for representativeness were large for both the game ($d_{\text{pre-post}} = 1.51$ and $d_{\text{pre-followup}} = 1.05$) and video ($d_{\text{pre-post}} = 1.80$ and $d_{\text{pre-followup}} = 1.09$). The game more effectively debiased participants than did the

video immediately, $F(1, 235) = 10.85$, $p < .01$, but was no more effective three months later, $F < 1$, $p = .37$.

Bias Knowledge

Training effectively improved bias knowledge immediately and three months later, $F(1, 236) = 506.52$, $p < .001$ and $F(1, 190) = 216.36$, $p < .001$, respectively. Bias knowledge increased for participants who played the game ($M_{pretest} = 35.89$, $M_{posttest} = 63.16$, $M_{follow\text{-}up} = 50.65$, $d_{pre\text{-}post} = .1.42$ and $d_{pre\text{-}followup} = 1.05$) and watched the video ($M_{pretest} = 39.03$, $M_{posttest} = 74.11$, $M_{follow\text{-}up} = 52.04$, $d_{pre\text{-}post} = 1.53$ and $d_{pre\text{-}follow\text{-}up} = 1.09$). The video more effectively taught participants to recognize and discriminate bias than did the game immediately, $F(1, 235) = 11.07$, $p < .001$, but was no more effective three months later, $F < 1$.

CONCLUSIONS AND RECOMMENDATIONS

People generally intend to make good decisions, which are in their own and society's best interest, but biases in judgment and decision making often lead them to make costly errors. More than 40 years of judgment and decision making research suggests feasible interventions to debias and improve decision making (Bhargava & Loewenstein, 2015; Fischhoff, 1982; Fox & Sitkin, 2015; Soll et al., in press). This research and its methods can be used to align incentives, present information, elicit choices, and educate people so they are able to make decisions in their best interest.

Debiasing interventions are not, by default, coercive. Decisions always have some underlying structure that may bias the process or the outcome. Presenting information in a manner in which options are easier to evaluate generally improves choices by making people better able to evaluate those options along the dimensions that are important to

them. Commuting ranks among the most unpleasant daily experiences (Kahneman, Krueger, Schkade, Schwarz, & Stone, 2004), for instance, but people are relatively insensitive to the duration of a prospective commute unless they are provided with a familiar comparison standard (Morewedge, Kassam, Hsee, & Caruso, 2009). For some decisions such as whether to be an organ donor, one option must be specified as the default even if one defers the decision. Selecting a default option that is beneficial for the decision maker or society can improve the public good while preserving freedom of choice (Camerer et al., 2003; Thaler & Sunstein, 2003; 2008). Furthermore, people actively seek out many kinds of debiasing interventions such as timesaving recommendation systems (Goldstein et al., 2008) and commitment devices to give them the willpower to make choices that are unappealing in the present but will benefit them more in the future (e.g., Thaler & Benartzi, 2004; Schwarz et al., 2014).

Debiasing interventions are not, by default, more costly than the status quo. New incentives do not have to impose a financial cost to taxpayers or decision makers. Social influence is an underutilized but powerful nonpecuniary motive for positive behavior change, for instance, that can produce significant reductions in environmental waste and energy consumption (Cialdini, 2003; Schultz, Nolan, Cialdini, Goldstein, & Griskevicius, 2007). Moreover, existing incentives are only effective if they motivate behavior as they were intended. If incentives are misaligned, misinterpreted, or poorly framed, they may be costly and ineffective or counterproductive.

Small changes in message framing and choice elicitation can produce debiasing effects for little additional cost. In two laboratory studies, simply framing an economic stimulus as a “bonus” rather than a “rebate” more than doubled how much of that

stimulus was spent (Epley, Mak, & Idson, 2006). In a field study run in the United Kingdom, adding a single sentence to late tax notices that truthfully stated the majority of UK citizens pay their taxes on time increased the clearance rate of late payers to 86% (£560 million out of £650 million owed), compared to a clearance rate of 57% the previous year (£290 million out of £510 million owed; Cialdini, Martin, & Goldstein, 2015).

Training interventions have an upfront production cost, but the marginal financial and temporal costs of training many additional people are minimal. The results of our experiments suggest that even a single training intervention, such as the games and videos we tested in this article, can have significant debiasing effects that persist across a variety of contexts affected by the same bias. Participants who played our games exhibited large reductions in cognitive bias immediately (-46.25% and -31.94%), which persisted at least 2 or 3 months later (-34.76% and -23.57%) in Experiments 1 and 2, respectively. Participants who watched the videos exhibited medium and large reductions immediately (-18.60% and -25.70%), which persisted at least 2 or 3 months later (-20.10% and -19.20%) in Experiments 1 and 2, respectively. The greater efficacy of the games than the videos suggest that personal feedback and practice increase the debiasing effects of training, but more research is needed to determine precisely why it was more effective. Most important, these results suggest that despite its rocky start (Fischhoff, 1982), training is a promising avenue through which to develop future debiasing interventions.

Decision research is in an exciting phase of expansion, increasing the basic research that identifies and elucidates biases while extending its reach by developing and

testing new practical interventions. Laboratory experiments provide a safe and inexpensive microcosm in which to uncover new biases, develop new theories, and test new interventions. Many are now testing successful laboratory interventions and their extensions in larger field experiments, such as randomized controlled trials, to determine which biases and interventions are most influential in particular contexts (Haynes, Service, Goldacre, & Torgerson, 2012). This work extends outside the ivory tower. Researchers have produced numerous successful collaborations with government and industry partners that have reduced waste and improved the health and finances of the public (e.g., Chapman et al., 2010; Mellers et al., 2014; Schultz et al., 2007; Schwartz et al., 2014; Thaler & Benartzi, 2004). Ad hoc collaborations and targeted programs, such as the development and testing of training inventions that we report, have been very successful (see also Mellers et al., 2014). Several countries have even established panels of behavioral scientists to develop interventions from within government, such the *Social and Behavioral Sciences Team* in the United States. Decision making is pervasive in professional and everyday life. Its study and improvement can contribute much to the public good.

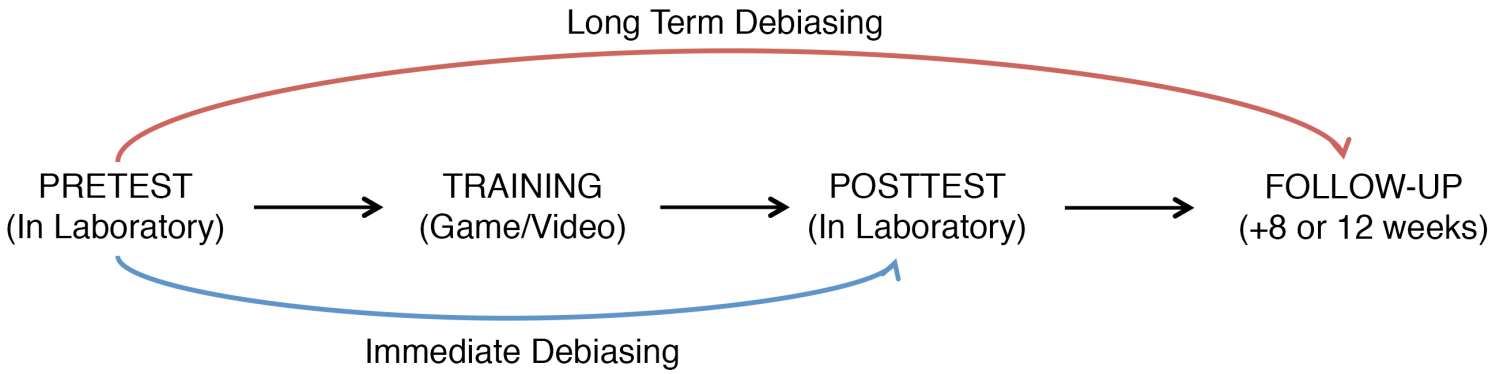

Figure 1. Immediate debiasing effects of training interventions (a game or video) were measured by comparing pretest and posttest scores of bias commission in a laboratory session. Long term debiasing effects of training interventions were measured in an online follow-up measuring bias commission 8 or 12 weeks later (Experiments 1 and 2, respectively).

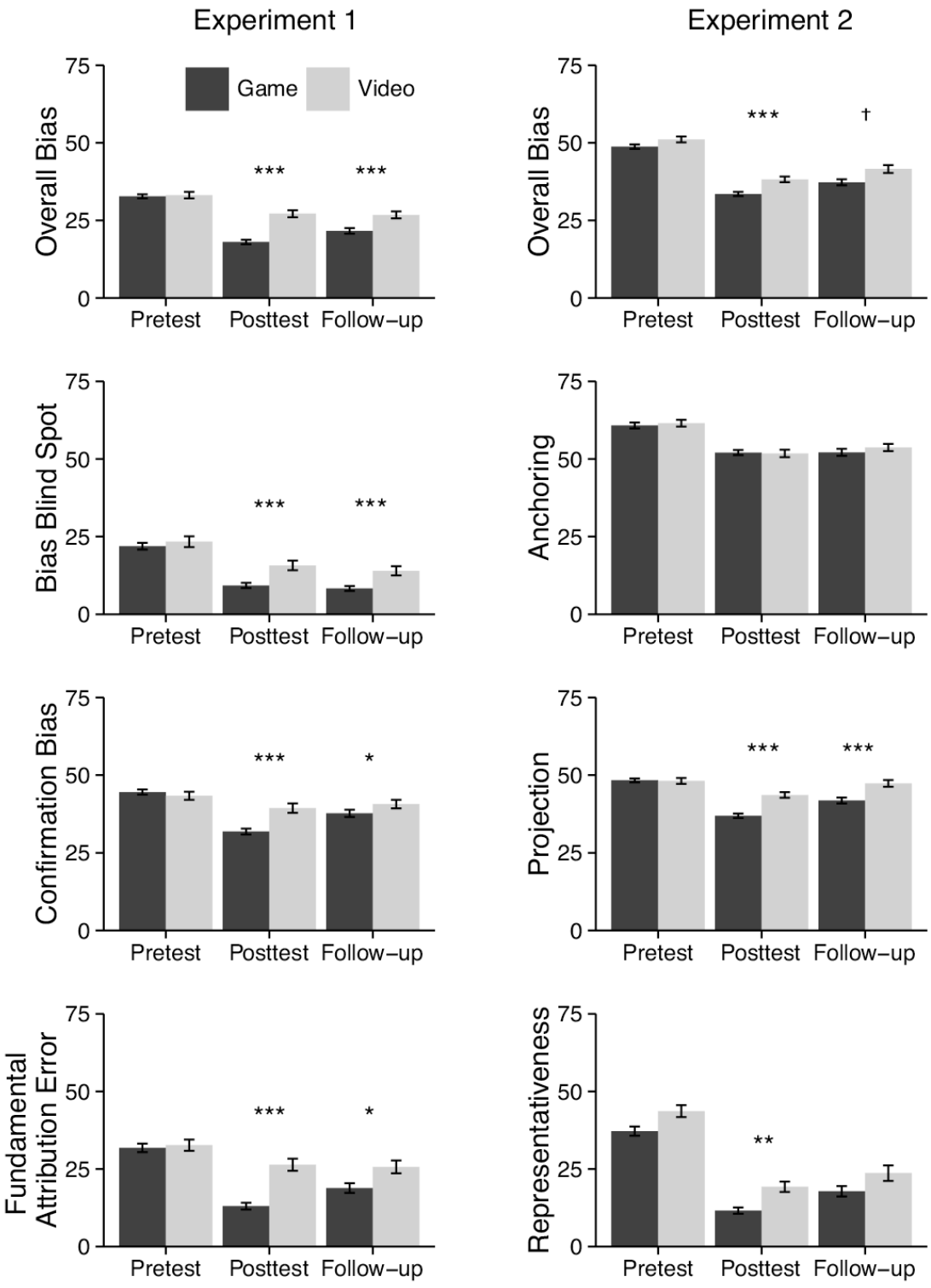
Experiment 1
Experiment 2
Game
Video
Overall Bias
Bias Blind Spot
Confirmation Bias
Fundamental Attribution Error
Anchoring
Projection
Representativeness
75
50
25
0
Pretest
Posttest
Follow–up
***
**
*
†

Figure 2. Bias commission by training intervention in Experiments 1 and 2. Left and right columns illustrate the mitigating effects of training on bias commission overall and for each of the three cognitive biases in Experiments 1 and 2, respectively. Scales range from 0-100; higher scores indicate more biased answers (95% CI). Both training interventions effectively debiased participants. Overall, the game more effectively debiased participants than did the video in Experiments 1 and 2. Symbols indicate statistically significant and marginally significant differences between game and video conditions at posttest and follow-up: $^{\dagger}p < .10$; * $p < .05$; ** $p < .01$; and ***$p < .001$.

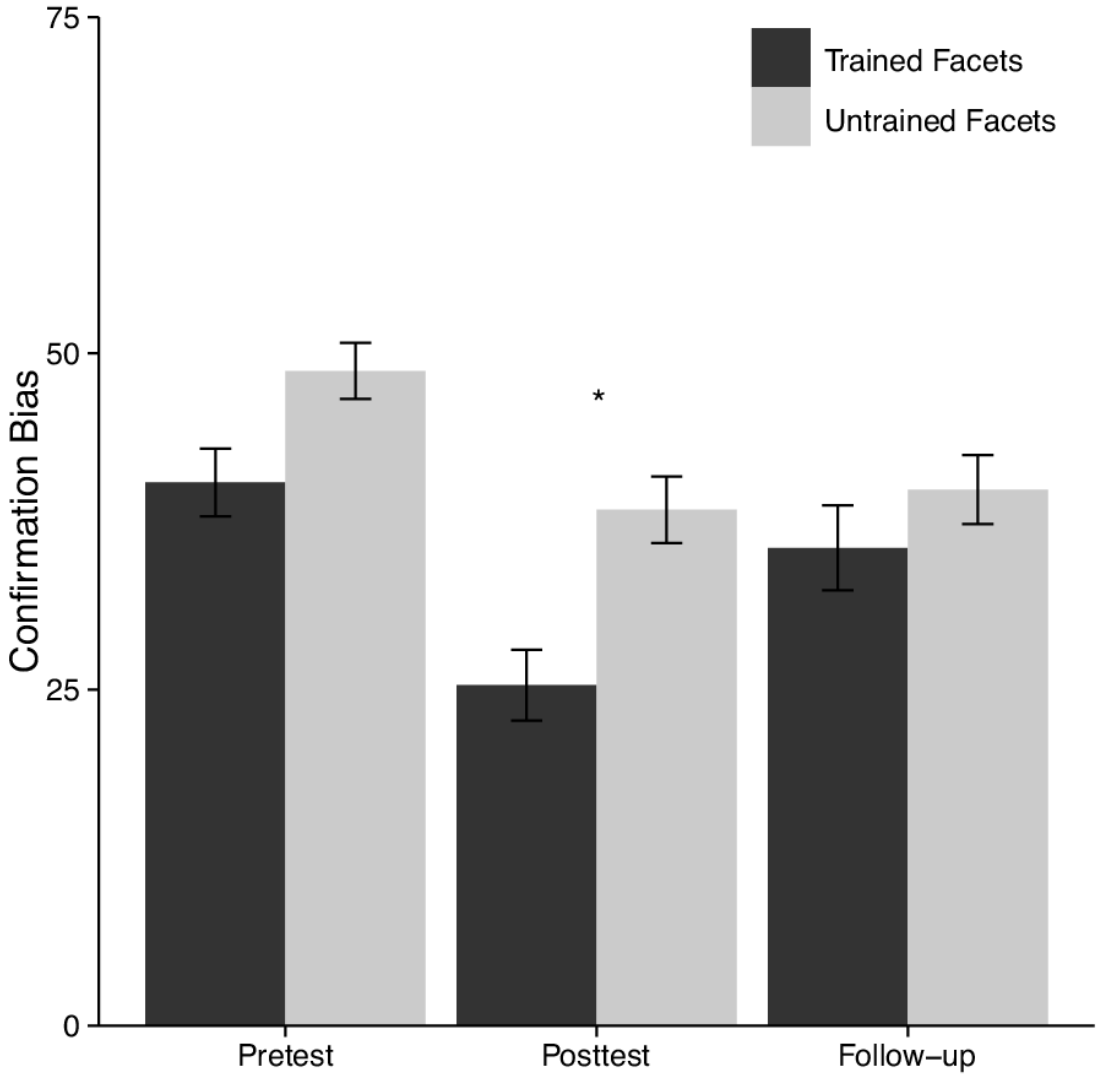

Figure 3. Debiasing effects of the game were observed for both trained and untrained facets of confirmation bias in Experiment 1, suggesting that debiasing effects of training generalized across domains. Scales range from 0-100, higher scores indicate more bias (95% CI). Asterisk indicates significant difference between trained and untrained facets of confirmation bias at posttest, controlling for pretest scores, * $p < .05$.

FOOTNOTES

1. There were four variants of the Experiment 1 game, including whether a game score or narrative examples were included or excluded in the AARs. Moreover, half of participants in the game condition played the full game, and half played only the first round. We did not observe a significant difference across these game and methodological variations in their reduction of overall bias at posttest and follow-up, Fs ≤ 2.29, $ps \geq .13$. In Experiment 2, all players completed the whole game, but there were four variants including whether hints or game scores were provided. We did not observe a significant difference across these variants in their reduction of overall bias at posttest and follow-up, all ts ≤ 1.77, $ps \geq .08$. In both experiments, we report the results collapsed across these variations.
2. Participants were excluded before analyses in Experiment 1 because they played early game prototypes ($n = 20$), experienced game crashes ($n = 3$) and server errors during scale administration ($n = 6$), or were unable to finish the laboratory session in 4 hours ($n = 6$). In addition, those who did not complete the follow-up test within 7 days of receiving notification were not included in follow-up analyses ($n = 47$). Participants were excluded before analyses in Experiment 2 because of game crashes ($n = 1$), experimenter or participant error ($n = 3$), or failed attention checks ($n = 27$). In addition, those who did not complete the follow-up within 7 days of receiving notification were not included in follow-up analyses ($n = 45$).

ACKNOWLEDGEMENTS

This work was supported by the Intelligence Advanced Research Projects Activity via the Air Force Research Laboratory contract number FA8650-11-C-7175. 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. We thank Marguerite Barton, Abigail Dawson, Sophie LeBrecht, Erin McCormick, Peter Mans, H. Lauren Min, Taylor Turrisi, and Shane Schweitzer for their assistance with the execution of this research.

Author to whom all correspondence should be addressed: Carey K. Morewedge, Associate Professor of Marketing; Boston University, Questrom School of Business, Rafik B. Hariri Building, 595 Commonwealth Ave., Boston, MA 02215; morewedg@bu.edu.

References

Acquisti, A., Brandimarte, L., & Loewenstein, G. (2015). Privacy and human behavior in the age of information. *Science*, *347*, 509-514.

Ariely, D., Gneezy, U., Loewenstein, G., & Mazar, N. (2009). Large stakes and big mistakes. *The Review of Economic Studies*, *76*, 451-469.

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

Arkes, H. R., & Blumer, C. (1985). The psychology of sunk cost. *Organizational behavior and human decision processes*, *35*, 124-140.

Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn?: A taxonomy for far transfer. *Psychological Bulletin*, *128*, 612-637.

Barton, M., Symborski, C., Quinn M., Morewedge, C. K., Kassam, K. S., Korris, J. H. (2015, May). *The use of theory in designing a serious game for the reduction of cognitive biases*. Digital Games Research Association Conference, Lüneberg, Germany.

Bhargava, S., & Loewenstein, G. (2015). Behavioral Economics and Public Policy 102: Beyond Nudging. *American Economic Review*, *105*, 396-401.

Bhargava, S., & Manoli, D. (in press). Psychological frictions and incomplete take-up of social benefits: Evidence from an IRS field experiment. *American Economic Review.*

Bruine de Bruin, W., Parker, A. M., & Fischhoff, B. (2007). Individual differences in adult decision-making competence. *Journal of Personality and Social Psychology*, *92*, 938-956.

Camerer, C., Issacharoff, S., Loewenstein, G., O'donoghue, T., & Rabin, M. (2003). Regulation for conservatives: Behavioral economics and the case for "asymmetric paternalism." *University of Pennsylvania Law Review*, 1211-1254.

Chapman, G. B., Li, M., Colby, H., & Yoon, H. (2010). Opting in vs opting out of influenza vaccination. *Journal of the American Medical Association*, *304*, 43-44.

Charness, G., & Gneezy, U. (2009). Incentives to exercise. *Econometrica*, *77*, 909-931.

Cialdini, R. B. (2003). Crafting normative messages to protect the environment. *Current Directions in Psychological Science, 12,* 105-109.

Cialdini, R. B., Martin, S. J., & Goldstein, N. J. (2015). Small behavioral science–informed changes can produce large policy-relevant effects. *Behavioral Science and Policy, 1*.

Cooper, J. R. (2005). *Curing analytic pathologies: Pathways to improved intelligence analysis.* Center for the Study of Intelligence: Central Intelligence Agency.

Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science*, *243*, 1668-1674.

Dhami, M. K., Schlottmann, A., & Waldmann, M. R. (2011). *Judgment and decision making as a skill: learning, development and evolution*. Cambridge, UK: Cambridge University Press.

Downs, J. S. (2014). Prescriptive scientific narratives for communicating usable science. *Proceedings of the National Academy of Sciences*, *111*, 13627-13633.

Downs, J. S., Loewenstein, G., & Wisdom, J. (2009). Strategies for promoting healthier food choices. *The American Economic Review,* 159-164.

Downs, J. S., & Shafir, E. (1999). Why some are perceived as more confident and more insecure, more reckless and more cautious, more trusting and more suspicious, than others: Enriched and impoverished options in social judgment. *Psychonomic Bulletin & Review*, *6*, 598-610.

Epley, N., Mak, D., & Idson, L. C. (2006). Bonus of rebate?: The impact of income framing on spending and saving. *Journal of Behavioral Decision Making, 19*, 213-227.

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,* 760-768.

Evans, J. S. B. (2003). In two minds: dual-process accounts of reasoning. *Trends in Cognitive Sciences*, *7*, 454-459.

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.

Fox, C. R., & Sitkin, S. B. (2015). Bridging the divide between behavioral science and policy. Behavioral Science & Policy, 1.

French, S. A. (2003). Pricing effects on food choices. *The Journal of Nutrition*, *133*, 841S-843S.

Garcia-Retamero, R., & Dhami, M. K. (2011). Pictures speak louder than numbers: on communicating medical risks to immigrants with limited non-native language proficiency. *Health Expectations, 14*, 46-57.

Garcia-Retamero, R., & Dhami, M. K. (2013). On avoiding framing effects in experienced decision makers. *The Quarterly Journal of Experimental Psychology, 66*, 829-842.

Gerber, A. S., & Rogers, T. (2009). Descriptive social norms and motivation to vote: Everybody's voting and so should you. *The Journal of Politics, 71,* 178-191.

Gilbert, D. T. (1998). Ordinary personology. In D. T. Gilbert, S. T. Fiske, & G. Lindsey (Eds.), *The handbook of social psychology* (Vol. 2, pp. 89-150).

Goldstein, D. G., Johnson, E. J., Herrmann, A., & Heitmann, M. (2008). Nudge your customers toward better choices. *Harvard Business Review*, *86*, 99-105.

Gneezy, U., Meier, S., & Rey-Biel, P. (2011). When and why incentives (don't) work to modify behavior. *The Journal of Economic Perspectives*, 191-209.

Gneezy, U., & Rustichini, A. (2000). A fine is a price. *Journal of Legal Studies, 29,* 1–17.

Haferkamp, N., Kraemer, N. C., Linehan, C., & Schembri, M. (2011). Training disaster communication by means of serious games in virtual environments. *Entertainment Computing*, *2*, 81-88.

Harvey, N. (2011). Learning judgment and decision making from feedback. In M. K. Dhami, A. Schlottmann, and M. R. Waldmann (eds.), *Judgment and decision making as a skill: Learning, development, and evolution* (pp. 199-226). Cambridge, UK: Cambridge University Press.

Haynes, L., Goldacre, B., & Torgerson, D. (2012). Test, learn, adapt: developing public policy with randomized controlled trials. *Cabinet Office-Behavioural Insights Team.*

Heuer, R. J. (1999). *Psychology of intelligence analysis.* Center for the Study of Intelligence: Central Intelligence Agency.

Hirt, E. R., & Markman, K. D. (1995). Multiple explanation: A consider-an-alternative strategy for debiasing judgments. *Journal of Personality and Social Psychology*, *69*, 1069-1086.

Huh, Y. E., Vosgerau, J., & Morewedge, C. K. (2014). Social Defaults: Observed Choices Become Choice Defaults. *Journal of Consumer Research*, *41*, 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 (2012). *Unbiasing your biases I*. Alexandria, VA: 522 Productions.

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

John, L. K., Loewenstein, G., Troxel, A. B., Norton, L., Fassbender, J. E., & Volpp, K. G. (2011). Financial incentives for extended weight loss: a randomized, controlled trial. *Journal of general internal medicine*, *26*, 621-626.

Johnson, E. J., & Goldstein, D. G. (2003). Do defaults save lives?. *Science*, *302*, 1338-1339.

Jones, E. E.; Harris, V. A. (1967). "The attribution of attitudes". *Journal of Experimental Social Psychology, 3*, 1–24.

Kahneman, D. (2003). A perspective on judgment and choice: mapping bounded rationality. *American psychologist*, *58*, 697-720.

Kahneman, D., Krueger, A. B., Schkade, D. A., Schwarz, N., & Stone, A. A. (2004). A survey method for characterizing daily life experience: The day reconstruction method. *Science, 306,* 1776-1780.

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgement of representativeness. *Cognitive Psychology, 3,* 430-454.

Kohler, D. J., Brenner, L., & Griffin, D. (2002). The calibration of expert judgment: Heuristics and biases beyond the laboratory. In T. Gilovich, D. Griffin, and D. Kahneman (Eds), *Heuristics and biases: The psychology of intuitive judgment* (pp. 686-715). New York, NY, US: Cambridge University Press.

Kopecky, J. J., McKneely, J. A., & Bos, N. D. (2015). Personal communication.

Krieger, L. H., & Fiske, S. T. (2006). Behavioral realism in employment discrimination law: Implicit bias and disparate treatment. *California Law Review*, 997-1062.

Larrick, R. P., Morgan, J. N., & Nisbett, R. E. (1990). Teaching the use of cost-benefit reasoning in everyday life. *Psychological Science*, *1*, 362-370.

Larrick, R. P., & Soll, J. B. (2008). The MPG Illusion. *Science*, *320*(5883), 1593-1594.

Lerner, J. S., & Tetlock, P. E. (1999). Accounting for the effects of accountability. *Psychological bulletin*, *125*, 255-275.

Levin, I. P., & Gaeth, G. J. (1988). How consumers are affected by the framing of attribute information before and after consuming the product. *Journal of Consumer Research, 15,* 374-378.

Lewandowsky, S., Ecker, U. K., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and its correction continued influence and successful debiasing. *Psychological Science in the Public Interest, 13*, 106-131.

Lilienfeld, S. O., Ammirati, R., & Landfield, K. (2009). Giving debiasing away: Can psychological research on correcting cognitive errors promote human welfare?. *Perspectives on Psychological Science*, *4*, 390-398.

Mazarr, M. J. (2007). The Iraq war and agenda setting. *Foreign Policy Analysis*, *3*, 1-23.

McNeil, B. J., Pauker, S. G., Sox Jr, H. C., & Tversky, A. (1982). On the elicitation of preferences for alternative therapies. *New England Journal of Medicine*, *306*, 1259-1262.

Mellers, B., Ungar, L., Baron, J., Ramos, J., Gurcay, B., Fincher, K., ... & Tetlock, P. E. (2014). Psychological strategies for winning a geopolitical forecasting tournament. *Psychological Science, 25,* 1106-1115.

Milkman, K. L., Chugh, D., & Bazerman, M. H. (2009). How can decision making be improved?. *Perspectives on Psychological Science*, *4*, 379-383.

Moore, D. A., Swift, S. A., Sharek, Z. S., & Gino, F. (2010). Correspondence bias in performance evaluation: Why grade inflation works. *Personality and Social Psychology Bulletin*, *36*, 843-852.

Morewedge, C. K., Kassam, K. S., Hsee, C. K., & Caruso, E. M. (2009). Duration sensitivity depends on stimulus familiarity. *Journal of Experimental Psychology: General, 138,* 177-186.

Morewedge, C. K., & Kahneman, D. (2010). Associative processes in intuitive judgment. *Trends in cognitive sciences*, *14*, 435-440.

Murphy, A. H. & Winkler, R. L. (1974). Subjective probability forecasting experiments in meteorology: Some preliminary results. *Bulletin of the American Meteorological Society, 55,* 1206-1216.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of general psychology*, *2*, 175-220.

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

Nisbett, R. E., & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgment.* Englewood Cliffs, NJ: Prentice Hall.

Parker, A. M., & Fischhoff, B. (2005). Decision-making competence: External validation through an individual-differences approach. *Journal of Behavioral Decision Making*, *18*, 1-27.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision maker.* Cambridge University Press.

Peters, E., & de Bruin, W. B. (2011). Aging and decision skills. In M. K. Dhami, A. Schlottmann, and M. R. Waldmann (eds.), *Judgment and decision making as a skill: Learning, development, and evolution* (pp. 113-140). Cambridge, UK: Cambridge University Press.

Phillips, J. K., Klein, G., & Sieck, W. R. (2004). Expertise in judgment and decision making: A case for training intuitive decision skills. *Blackwell handbook of judgment and decision making*, 297-315.

Reeves, L., & Weisberg, R. W. (1994). The role of content and abstract information in analogical transfer. *Psychological Bulletin, 115*, 381-400.

Robbins, J. M., & Krueger, J. I. (2005). Social projection to ingroups and outgroups: A review and meta-analysis. *Personality and Social Psychology Review*, *9*, 32-47.

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*, 279-301.

Schultz, P. W., Nolan, J. M., Cialdini, R. B., Goldstein, N. J., & Griskevicius, V. (2007). The constructive, destructive, and reconstructive power of social norms. *Psychological Science, 18*, 429-434.

Schwartz, J., Mochon, D., Wyper, L., Maroba, J., Patel, D., & Ariely, D. (2014). Healthier by precommitment. *Psychological Science*, *25*, 538-546.

Schwitzgebel, E., & Cushman, F. (2012). Expertise in moral reasoning? Order effects on moral judgment in professional philosophers and non-philosophers. *Mind & Language*, *27*, 135-153.

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*.

Shefrin, H., & Statman, M. (1985). The disposition to sell winners to early and ride losers too long: Theory and evidence. *Journal of Finance, 40,* 777-790.

Silberman, L. H., Robb, C. S., Levin, R. C., McCain, J., Rowan, H. S., Slocombe, W. B., Studeman, W. O., Wald, P. M., Vest, C. M., & Cutler, L. (March 31, 2005). *The Commission on the Intelligence Capabilities of the United States Regarding Weapons of Mass Destruction: Report to the President of the United States.*

Soll, J., Milkman, K., & Payne, J. (in press). A user's guide to debiasing. In G. Keren and G. Wu (eds), *Wiley-Blackwell handbook of judgment and decision making*. New York, NY: Blackwell.

Sliney, A., & Murphy, D. (2008, February 10-15). *JDoc: A serious game for medical learning*. Proceedings of the First International Conference on Advances in Computer-Human Interaction (ACHI-2008), Sainte Luce, Martinique. doi: 10.1109/ACHI.2008.50

Sloman, S. A. (1996). The empirical case for two systems of reasoning.*Psychological bulletin*, *119*, 3-22.

Snyder, M., & Swann, W. B. (1978). Hypothesis-testing processes in social interaction. *Journal of Personality and Social Psychology*, *36*, 1202-1212.

Symborski, C., Barton, M., Quinn M., Morewedge, C. K., Kassam, K., Korris, J. (2014, December). *Missing: A serious game for the mitigation of cognitive bias*. Interservice/Industry Training, Simulation and Education Conference, Orlando, FL.

Thaler, R. H., & Benartzi, S. (2004). Save more tomorrow™: Using behavioral economics to increase employee saving. *Journal of political Economy*, *112*, S164-S187.

Thaler, R. H., & Sunstein, C. R. (2003). Libertarian paternalism. *American Economic Review*, 175-179.

Thaler, R.H., & Sunstein, C.R. (2008). *Nudge*. New Haven, CT: Yale University Press.

Tschirgi, J. E. (1980). Sensible reasoning: A hypothesis about hypotheses. *Child Development*, *51,* 1-10.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*, 1124-1131.

Volpp, K. G., Loewenstein, G., Troxel, A. B., Doshi, J., Price, M., Laskin, M., & Kimmel, S. E. (2008). A test of financial incentives to improve warfarin adherence. *BMC Health Services Research*, *8*, 272.

Volpp, K. G., Troxel, A. B., Pauly, M. V., Glick, H. A., Puig, A., Asch, D. A., . . . Audrain-McGovern, J. (2009). A randomized, controlled trial of financial incentives for smoking cessation. *New England Journal of Medicine*, *360*, 699–709.

Wagenaar, W. A., & Keren, G. B. (1986). Does the expert know? The reliability of predictions and confidence ratings of experts. In *Intelligent decision support in process environments* (pp. 87-103). Springer: Berlin Heidelberg.

Wason, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology*, *12*, 129-140.

Wason, P. C. (1968). Reasoning about a rule. *The Quarterly journal of experimental psychology*, *20*(3), 273-281.

Willingham, D. T. (2008). Critical thinking: Why is it so hard to teach?. *Arts Education Policy Review*, *109*, 21-32.

Wilson, T. D., & Brekke, N. (1994). Mental contamination and mental correction: unwanted influences on judgments and evaluations. *Psychological Bulletin*, *116*, 117-142.

SUPPLEMENTAL MATERIALS

EXPERIMENT 1: SCALE DEVELOPMENT

For each bias, we conducted a literature review to identify canonical questions, paradigms, and generated similar additional items (approximately 200 in total). BBS questions were developed following the question format of Scopelliti and colleagues (2015). FAE questions were based on the attitude attribution, quizmaster, silent interview, and moral attribution paradigms (Gawronski, 2004). CB questions were developed based on six paradigms: Wason's (1960) card selection task, Wason's (1968) triplets task; Tschirgi's (1980) cause identification paradigm, Snyder and Swann's (1978) trait hypothesis testing paradigm, an enriched versus impoverished profiles choice paradigm (Downs & Shafir, 1999), and a judgment of covariation paradigm (Nisbett & Ross, 1980). Three interchangeable versions (i.e., subscales) were created for each scale, so that each participant would see different questions at pretest (before training), posttest (immediately after training), and follow-up (8 weeks after training).

One sample of 288 Amazon Mechanical Turk (AMT) workers answered all FAE and BBS items. A separate sample of 310 AMT workers answered all CB items. We performed scale purification using an iterative procedure. In order to ensure that three valid and interchangeable versions of each scale were developed, questions with low item-total correlations were removed until random sampling suggested that a subsample of one third of the items on each bias scale would achieve a minimum of $\alpha \geq .7$ reliability at least 95% of the time. This purification resulted in a 27-item BBS scale, a 45-item FAE scale, two 9-item scales based on Wason (1960, 1968), a 12-item scale based on Tschirgi

(1980), and three 18-item scales based on Snyder and Swann (1978), Downs and Shafir (1999), and Nisbett and Ross (1980). Exploratory factor analyses of the purified scales indicated a unidimensional structure for each scale, with average variance explained = 36%. All items correlated positively with their respective factor, with an average minimum $r = .41$.

Seven to 11 days after completing the full scales, 305 participants completed purified versions of the same scales. Responses indicated high test-retest reliability and stability over time, $M_r = .79$. We divided each scale into three interchangeable subscales by iteratively selecting the three items with the highest average correlations and placing them into separate subscales, all subscale α's $\geq .65$. Items were divided among subscales so that subscales were maximally similar.

Item scoring logic varied due to their different formats. All item scores varied between 0 and 1, with 1 indicating greater bias (i.e., choosing confirming answers, making dispositional attributions, indicating less susceptibility to bias than one's peers). We calculated subscale scores by summing all individual items (i.e., all BBS items, FAE items, or CB items) and transforming totals into a score ranging from 0 (no biased answers) to 100 (all answers biased).

Bias knowledge questions had two forms. Recognition questions described an instance in which one of the three biases was committed and required participants to identify the bias in a free recall format. Discrimination questions described an instance of bias and tested its identification in a multiple-choice format. The final questionnaires contained 24 questions with satisfactory face validity (12 for recognition and 12 for discrimination), equally divided among the three biases.

EXPERIMENT 2: SCALE DEVELOPMENT

Three interchangeable subscales were created for each scale, so that each participant would see different questions at pretest (before training), posttest (immediately after training), and follow-up (8 weeks after training). For each bias, we conducted a literature review to identify canonical questions, paradigms, and generated similar additional items (423 in total). Anchoring questions used self-generated or experimenter provided anchors that were relevant or irrelevant (Strack & Mussweiler, 1997; Tversky & Kahneman, 1974; Simmons, Nelson, & LeBoeuf, 2010). Projection questions were developed from three bias facets: the false consensus effect (Ross, Greene, & House, 1976), attributive similarity (Holmes, 1968; Kreuger & Stanke, 2001), and the curse of knowledge (Birch & Bloom, 2007). The curse of knowledge dimension was not included in the final instrument based on factor analyses suggesting its exclusion. Representativeness questions were based on conjunction fallacy, base rate neglect, gambler's fallacy, perceptions of random sequences, and sample size neglect paradigms (Tversky & Kahneman, 1974).

After an initial purification stage, three samples of AMT workers ($N = 624$) completed the scales. Purification resulted in a 54-item anchoring scale, a 69-item projection scale, and a 78-item representativeness scale. Questions were then split into three interchangeable subscales for each bias. All the subset scales had acceptable internal consistency, $M_\alpha = .68$, and test re-test reliability, $M_r = .61$. Item and subscale scoring logic followed the procedure used in Experiment 1.

All bias knowledge questions for Experiment 2 were multiple choice discrimination questions. The final questionnaire contained 21 questions with satisfactory

face validity, divided into subscales with 7 questions each. Knowledge scales were scored on a 0-100 scale with higher scores indicating greater knowledge.

References

Birch, S. A., & Bloom, P. (2007). The curse of knowledge in reasoning about false beliefs. *Psychological Science*, *18*, 382-386.

Gawronski, B. (2004). Theory-based bias correction in dispositional inference: The fundamental attribution error is dead, long live the correspondence bias. *European review of social psychology*, *15*(1), 183-217.

Downs, J. S., & Shafir, E. (1999). Why some are perceived as more confident and more insecure, more reckless and more cautious, more trusting and more suspicious, than others: Enriched and impoverished options in social judgment. *Psychonomic bulletin & review*, *6*, 598-610.

Holmes, D. S. (1968). Dimensions of projection. *Psychological Bulletin*, *69*(4), 248-268.

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.

Nisbett, R. E., & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgment.* Englewood Cliffs, NJ: Prentice Hall.

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*.

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*, 917-932.

Snyder, M., & Swann, W. B. (1978). Hypothesis-testing processes in social interaction. *Journal of Personality and Social Psychology*, *36*(11), 1202-1212.

Strack, F., & Mussweiler, T. (1997). Explaining the enigmatic anchoring effect: Mechanisms of selective accessibility. *Journal of Personality and Social Psychology*, *73*(3), 437-446.

Tschirgi, J. E. (1980). Sensible reasoning: A hypothesis about hypotheses. *Child development*, 1-10.

Wason, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly journal of experimental psychology*, *12*(3), 129-140.

Wason, P. C. (1968). Reasoning about a rule. *The Quarterly journal of experimental psychology*, *20*(3), 273-281.