Game-based Training to Mitigate Three Forms of Cognitive Bias

Benjamin A. Clegg
Colorado State University
Fort Collins, CO
benjamin.clegg@colostate.edu

Rosa Mikeal Martey
Colorado State University
Fort Collins, CO
rosa.martey@colostate.edu

Jennifer Stromer-Galley
Syracuse University
Syracuse, NY
jstromer@syr.edu

Kate Kenski
University of Arizona
Tucson, AZ
kkenski@email.arizona.edu

Tobi Saulnier
1st Playable Productions
Troy, NY
tobi@1stplayable.com

James E. Folkestad
Colorado State University
Fort Collins, CO
James.Folkestad@colostate.edu

Elizabeth McLaren
1st Playable Productions
Troy, NY
elizabeth@1stplayable.com

Adrienne Shaw
Temple University
Philadelphia, PA
adrienneshaw13@gmail.com

Joanna E. Lewis
University of Central Florida
Orland, FL
joanna.lewis@knights.ucf.edu

John D. Patterson
Binghamton University
Binghamton, NY
jpatter4@binghamton.edu

Tomek Strzalkowski
University at Albany
Albany, NY
tomek@albany.edu

ABSTRACT

Cognitive biases are systematic errors that result from reliance on heuristics in decision-making. Such biases are typically automatic and unconscious influences on behavior, and can occur in a wide range of situations and contexts. Cognitive biases are generally resistant to mitigation training. This project adopted a novel approach to develop computer game-based training to attempt to mitigate three forms of cognitive bias: fundamental attribution error, the tendency to assume dispositional rather than situational influences account for behavior of others; confirmation bias, the tendency to seek and remember information that matches or supports one's view; and bias blind spot, the tendency to regard one's own decisions as being free from cognitive bias, even where one can recognize that bias in others. Participants were randomly assigned to play the training game once, or repeated twice with a 7-10 day delay between sessions (mean duration first play=43 minutes; second play=34 minutes), or to a control condition that employed a 30-minute professionally developed training video. Effects of training were measured on external questionnaire-based items, both immediately post-exposure, and at an 8-week retention interval. The game was intended to develop conceptual understanding of these biases, and recognition of circumstances within which they might occur. Using notional "tools" presented within the game, participants learned and practiced strategies to avoid decision-making influenced by the cognitive biases. Results showed that the training game successfully reduced bias on the assessment instrument, and outperformed the video both immediately post-training and at the retention test. Repetition of the training game did not further advantage immediate post-test performance but significantly improved retention. Validation of the key findings was confirmed by an independent group who used the training game with their own novel bias assessment instruments (to which the researchers and game-developers had no access or content information).

ABOUT THE AUTHORS

Dr. **Benjamin Clegg** is a Professor of Cognitive Psychology in the Department of Psychology at Colorado State University.

Dr. **Rosa Mikeal Martey** is an Associate Professor in the Department of Journalism and Communication at Colorado State University.

Dr. **Jennifer Stromer-Galley** is an Associate Professor in the School of Information Studies and affiliated faculty member with the Department of Political Science in the Maxwell School at Syracuse University.

Dr. **Kate Kenski** is an Associate Professor in the Department of Communication and School of Government & Public Policy at the University of Arizona.
Dr. **Emilie "Tobi" Saulnier** is the founder and CEO of game development studio 1st Playable Productions.
Dr. **James Folkestad** is currently an Associate Professor in the School of Education and the Director of the Analytics for Learning and Teaching (ALT) Center at Colorado State University.
Elizabeth McLaren is a Lead Artist and Production Coordinator at 1st Playable Productions.
Dr. **Adrienne Shaw** is an Assistant Professor in the Media Studies and Production Department and School of Media and Communication at Temple University.
Joanna Lewis is a graduate student in the Department of Psychology at University of Central Florida and a NSF Graduate Fellow.
JD Patterson is a Graduate Student in Cognitive Psychology at Binghamton University.
Dr. **Tomek Strzalkowski** is a Professor of Computer Science and Director of ILS Institute at the University at Albany.

Game-based Training to Mitigate Three Forms of Cognitive Bias

Benjamin A. Clegg
Colorado State University
Fort Collins, CO
benjamin.clegg@colostate.edu

Rosa Mikeal Martey
Colorado State University
Fort Collins, CO
rosa.martey@colostate.edu

Jennifer Stromer-Galley
Syracuse University
Syracuse, NY
jstromer@syr.edu

Kate Kenski
University of Arizona
Tucson, AZ
kkenski@email.arizona.edu

Tobi Saulnier
1st Playable Productions
Troy, NY
tobi@1stplayable.com

James E. Folkestad
Colorado State University
Fort Collins, CO
James.Folkestad@colostate.edu

Elizabeth McLaren
1st Playable Productions
Troy, NY
elizabeth@1stplayable.com

Adrienne Shaw
Temple University
Philadelphia, PA
adrienneshaw13@gmail.com

Joanna E. Lewis
University of Central Florida
Orland, FL
joanna.lewis@knights.ucf.edu

John D. Patterson
Binghamton University
Binghamton, NY
jpatter4@binghamton.edu

Tomek Strzalkowski
University at Albany
Albany, NY
tomek@albany.edu

Although our tendency to use heuristics can lead to efficient and often accurate decisions (Gigerenzer & Brighton, 2009), cognitive biases are systematic errors that may result from these heuristics (Tversky & Kahneman, 1974). Such biases have been considered automatic and unconscious influences on behavior (Kahneman, 2011). Cognitive biases can impair optimal decision-making in a wide range of tasks within business, medical, legal, and military settings (e.g., Nickerson, 1998). Domain knowledge does not overcome the effects of cognitive biases and may increase the effects (Reyna, Chick, Corbin, & Hsia, 2014). In addition, these forms of bias are not systematically related to personality attributes (West, Meserve, & Stanovich, 2012). Thus issues arising from cognitive bias exist in many real-world contexts and are not likely to be eliminated through careful personnel selection or increased expertise.

A variety of approaches to debiasing exist (Larrick, 2004), including development of decision support systems (Arnott, 2006), checklist procedures for decision-making within organizations (Kahneman, Lovallo, & Sibony, 2011), and modifications to the environment (Klayman & Brown, 1993). However, such methods typically invoke external interventions into the decision making process. Cognitive biases are generally expected to be resistant to generic mitigation training at the level of an individual (Fischhoff, 1982; Kahneman, 2003), and such training can actually exacerbate rather than reduce the bias (e.g., Sanna, Schwartz, & Stocker, 2002). This project adopted a novel approach to develop computer game-based training to attempt to teach individuals to mitigate three forms of cognitive bias. It is important to note that we do not claim that the use of video games inherently produces benefits over other forms of training and learning. Rather such a platform is one clear avenue to incorporate an array of desirable elements into the training experience. Whether video games are better than other forms of training is a separate question for future research. However, the current study does employ professionally developed video to serve as a control comparison for benefits that can be derived from a more passive form of learning of comparable underlying information.

The game used in this project aimed to train recognition and mitigation of three types of cognitive biases. Fundamental attribution error (FAE; also called Correspondence Bias) is the tendency to assume dispositional rather than situational influences account for the behavior of others (see Gilbert & Malone, 1995). Confirmation bias (CB) is the tendency to seek and remember information that matches or supports one's view (see Nickerson, 1998). Bias blind spot (BBS) is the tendency to regard one's own decisions as being free from cognitive bias, even when one can recognize that bias in others (see Pronin, 2007).

A key element of our approach to training was application of Transformative Learning Theory (TLT; Mezirow, 1990) using the Observe-Orient-Decide-Act Loop (OODA; Boyd, 1995) as a core aspect of game design. This approach leverages the dynamic and complex learning environment offered by digital games and allows us to address bias mitigation better than other techniques, such as "drill-and-skill". Transformative Learning Theory aims for "perspective transformation" along three dimensions: psychological (understanding of the self), convictional (belief systems), and behavioral (actions). TLT suggests that changes in these dimensions can result from both instrumental learning, or learning through task-oriented problem solving, and communicative learning, or understanding others' values, ideas, feelings, and/or moral decisions. We apply TLT to bias mitigation using an evolutionary learning and decision-making model, the OODA Loop, which posits that individuals: observe an evolving situation; orient assessments via personal, cultural, and other filters; make decisions based on filtered observations; take actions based on those decisions. The *orient* stage, during which individuals assess their own perspectives and assumptions, is of particular importance to bias mitigation. Both the OODA and TLT emphasize cycles of learning that lead people towards more self-reflection, resulting in enhanced assimilation of new concepts that can be translated into actions (Mezirow, 1990).

Additional potential strengths for the use of computer-games for training include that they allow for experiential learning, can provide a structured learning environment, an engaging and active learning situation, and the potential to tailor learning and instruction based on performance. The game format also readily allows for the inclusion of a range of phenomena known to enhance learning. These included spaced practice from the separation of encounters with related material across the game (*spaced practice,* e.g., Melton, 1970), quiz questions and other types of testing effects forcing retrieval of information that is advantageous compared to restudying information (*testing effect,* e.g., Roediger & Karpicke, 2006), and the active construction of information rather than passive reading (*generation effect,* e.g., Slamecka & Graf, 1978). One further element of the design used was to teach a set of techniques tied to notional "tools" within the game. For confirmation bias, the game provided a tool that was a set of scales and encouraged participants to seek a balance of confirming and disconfirming evidence. For fundamental attribution error, players used a set of goggles to use to seek potential situational explanations rather than accepting a dispositional assumption. For bias blind spot, participants received a warning sign to remind them they were always "blind to their own biases," whether they felt they were acting with bias or not. Biases and mitigation techniques were also explained using the tools as reference points, and integrating icons for these tools was intended to enhance recall.

In order to demonstrate the effectiveness of training, testing incorporated two critical elements. First examination of learning and mitigation was conducted outside of the game environment. The need for transfer from training to other settings (Holding, 1991) should be a self-evident requirement of an effective process. Within the current study we used an external survey measuring each bias to examine this transfer to outside of the context of the training. Second, the inability of trainees to retain newly learned knowledge and skills is well-documented in the psychological literature (e.g., Annett, 1979). Moreover, some forms of instruction and training maximize immediate performance at the end of acquisition, whereas learning is best considered from the retention of knowledge (see Schmidt & Bjork, 1992). Given this separation between what causes one to learn the most within the training experience and what you actually remember in the future, knowledge and mitigation were examined both immediately after training and at 8 weeks after training (with no refresher training between acquisition and retention). Our independent variable manipulated a spaced repetition of game play in contrast to a single encounter. Such a repetition, especially when spaced by 10-20% of the required retention interval (Cepeda et al., 2008), has been shown to enhance retention with other types of material.

The current study therefore sought to address two core questions: (1) Can bias knowledge and mitigation for each of the three biases be effectively trained within the novel computer-game that was developed? (2) Does spaced repetition of the game provide advantages to retention of any learning obtained?

METHODS

Stimuli

The game was designed as a Flash-based casual puzzle-style game in which players navigated a third-person view of their avatar through a series of rooms encountering puzzles that teach about the biases and how to mitigate them. Players also received infographics before each room to define and describe biases and mitigation strategies, and short interactive quizzes followed each room to further reinforce the lesson of each room (See Figure 1 for examples of elements from the game). A help system was available that provided additional instructions for players who were

Figure 1. Example scenes from the game. Clockwise from upper left: An infographic teaching key information. A room with a puzzle for the player to solve that illustrates the bias and its mitigation. An exit sequence requiring generation of core material. A quiz question testing core knowledge.

unable to advance, and hints occasionally appeared at the bottom of the screen to remind players of the task or to provide key information needed to make progress in the room.

The game consisted of nine modular rooms. Each room featured a content introduction for the relevant cognitive bias, a series of activities to be conducted within the room, and a door puzzle requiring application of core content to exit the room. Following room exit, participants encountered a summary and testing transition area before moving on to the next room. Room 1 featured an introduction to cognitive biases, the game's cover story, and taught an initial lesson of bias blind spot. Room 2 introduced confirmation bias, with players learning about confirming and disconfirming evidence. Room 3 featured fundamental attribution error and the idea of actively seeking all possible explanations for a situation, action, or event. Room 4 provided a puzzle involving confirmation bias. Room 5 taught mitigation of fundamental attribution error. Room 6 introduced balancing confirming and disconfirming evidence to mitigate confirmation bias. Room 7 examined the fundamental attribution error mitigation strategy. Room 8 examined bias blind spot and its mitigation strategy. Room 9 offered a final overall summary, and tested and reviewed core knowledge from the training.

To determine the efficacy of the training within the game a control condition with a training video was employed. The training video supplied to us had been written with input from subject matter experts, was professionally acted and produced, and provided the same core information as the video game with regard to the definitions and general bias mitigation strategies. The video provided a series of vignettes in which people exhibited cognitive biases. Those scenarios were then dissected by a "professor" who explained the interaction, the causes of bias and ways to mitigate the bias. The video employed humor and real-world relevant examples to maintain interest and engagement.

Procedure

Participants were recruited from college classes and psychology pools at three universities in the United States. Participants signed up on a central website and were assigned a unique identifier. Participants were then scheduled to come to a campus computer lab for 2-3 hours. After filling out a consent form, participants answered a questionnaire (mean duration=32 minutes, SD=8.3) on the computer delivered via the survey software Qualtrics. The items included university, internet/video game experience, FAE, CB, BBS, personal bias awareness, social conformity, Need for Cognition, gender roles, Big Five personality assessment, and demographics. Participants were then randomly assigned to one of three conditions: a single game play session ("Single"), two game play sessions ("Repeat"), or Training Video (mean duration of the first play of game= 43 minutes, second play=34 minutes; video=30 minutes). Immediately after completing the game/video, participants answered a post-session questionnaire on the computer using the survey software Qualtrics (mean duration=17, SD=5.9). Items on that questionnaire asked the following: university, recall of multimedia experience, transportation, active interpersonal engagement, social richness, social realism, mental immersion, general engagement, usability (ease, attention, stimulation, likability) perceptions, FAE, CB, BBS, personal bias awareness, and bias identification. A third survey was distributed via email and administered using Qualtrics software approximately eight weeks later for each participant (mean duration=34 minutes, SD=34). Questions included transportability (personality trait), recall of condition assignment, engagement (short version), usability (short version), learning perceptions, FAE, CB, BBS, personal bias awareness, and bias identification. Within this paper we focus only on the results of the bias knowledge and mitigation items.

Participants

Three hundred thirty-five subjects recruited for partial, optional course credit were included in our final dataset. The sample was majority female (67% female; 33% male), and 74% were between the ages of 18 and 20. One hundred ninety nine people participated in the retention study (a 59% response rate). They were paid a $20 gift card for completing this final survey.

Measures

A measure of recognition and identification of the biases, and individual scales to measure each of the cognitive biases -- FAE, CB, and BBS that were the focus of study -- were developed. As described below, the bias measures were scaled -100 to +100 with 0 indicating no bias in either direction (e.g. for FAE, -100 situational bias and +100 dispositional bias).

Bias Recognition. Bias recognition was assessed with six items on the pre- and post-survey, two items for each bias. Items were definitional statements, such as "believing you are fundamentally less susceptible to biased thinking than most people." Participants were asked to select the correct corresponding bias from a list of five (the list included distractor answers such as "Apathy Bias"). Recognition was scored by taking the sum of correct answers.

Fundamental Attribution Error. This measure presented participants with ten brief scenarios such as "One of your peers receives an 'A' in a course that has a reputation for being hard. The best explanation for this student's grade is that the student is smart. To what extent do you agree or disagree with this assessment?" Scenarios were based in range of topics areas such as donating money, romantic relationships, and driving. Similar to the rating scales used to assess attribution in Jones and Harris (1967) and Tetlock (1985), responses were gathered via sliders on a scale of 1 (*disagree*) to 7 (*agree*). The resulting rating was assessed relative to the center of the scale that was shown in other work to reflect a rating of agreement assigned to good situational account of these items. Hence in these answers a rating of 4 would reflect a dispositional account given no greater weighting than a situational one might receive. The measure was scored as follows: Scale Mean (range: 1 to 7)=Mean of 10 ratings. FAE Score (range: -100 to 100) = ((Scale Mean – 4) / 3) x 100.

Confirmation Bias. To examine CB, we tested how participants weighted information in relation to a known hypothesis (CB), based on the paradigm developed by Cook and Smallman (2008). For the CB measure used here**,** participants were presented with seven scenarios such as "You are considering taking a trip to the country Calzycoah. You don't speak the language, but you will not have to pay for airfare to this vacation spot." Then some scenarios provided a hypothesis, others asked participants to decide between two possible hypotheses (Taking, or not taking, the trip for the example above). Participants then rated six pieces of evidence for the importance of "asking the following questions in evaluating your selected decision." Three options represented confirming evidence and three disconfirming evidence. Answers used sliders on a scale of 1 (*unimportant*) to 7 (*extremely important*). The resulting ratings were assessed for the weight given to confirming evidence relative to the disconfirming evidence. The procedures used to calculate CB items (two sets of ratings for each of 7 questions, 3 confirming and 3 disconfirming items) were: Question Mean (two sets, range: 1 to 7) = Mean of 3 items per question. Adjusted Question Mean (two sets, range: 0 to 100) = ((Question Mean – 1) / 6) x 100. Scale Mean (two sets, range: -100 to 100) = Mean of 7 Adjusted Question Means. CB Score (range: -100 to 100) = (Scale Mean confirming items – Scale Mean disconfirming items).

Bias Blind Spot. BBS was measured with a variant of the method employed by Pronin, Lin & Ross (2002). Participants were first asked to rate themselves compared to an "average student at their institution" on seven different positive characteristics and traits. A second set of items then informed them of the Illusion of Superiority (IS), the tendency to rate oneself as above average on these types of dimensions. They were then asked, "To what extent do you believe that you showed this tendency when you rated your [intelligence] on a previous question?" This was followed with: "To what extent do you believe that the average student from your university would show this tendency if he or she rated his or her [intelligence]?" Answers are on sliders from 1 (*not at all*) to 9 (*very much*). Bias blind spot would be indicated when ratings for questions concern the self received lower values for susceptibility to the bias than an average student. BBS were scored as follows: Scale Mean (two sets, range: 1 to 9) = Mean of 7 ratings. Centered Score (range: -100 to 100) = ((5 - Scale Mean) / 4) x 100. BBS Score (range: -100 to 100) = (Centered Score self-ratings - Centered Score other-ratings) / 2.

Cronbach's alpha reliabilities were obtained for the three types of novel bias measures at pre-test (FAE=.68; CB=.59; BBS=.59), immediate post-test (FAE=.91; CB=.67; BBS=.66), and retention (FAE=.89; CB=.73;

BBS=.88). Pre-test scores also showed no evidence of correlations between the scales (all r<.1, p>.05), offering some evidence of discriminant validity.

Twenty participants who were not part of the main experiment were given a pre- and post-test on the bias measures with no feedback between administrations. The two attempts at the questions were separated by an intervening distractor (mathematical problems) task. Correlations between pre- and post-test suggest sufficient reliability for all measures of bias across the tests (test 1 to test 2 correlations, FAE: r=.94; CB: r=.69; BBS: r=.72). In addition the test-retest data also were used to examine whether practice effects occurred for the bias measures (FAE: mean (sd) test 1=10.4 (24.2), test 2=12.5 (25.6), $t(19)=1.10$, $p>.25$; CB: test 1=24.4 (13.8), test 2=24.7 (11.1), $t(19)=0.07$, $p>.90$; BBS: test 1=13.2, test 2 =12.4 (12.3), $t(19)=0.41$, $p>.65$) . Thus all scores remained stable across the two administrations of the instruments, with no evidence of significant decreases in bias from exposure to the bias questions alone.

Thirty five participants who were not part of the main study were used to examine whether pre-test exposure to the bias measures changed subsequent learning from the game (and hence post-test performance). There were no significant differences between the groups (pre + post-test vs. post-test only) on any of the bias measures at post-test (CB: $t(33)=0.03$, $p>.97$; FAE: $t(33)=-1.16$, $p>.25$; BBS: $t(33)=0.61$, $p>.54$). These findings show no evidence to suggest that the use of a pre-test changes the nature of learning to an extent that would influence the findings within the core study.

Design

Participants were randomly allocated to one of 3 conditions, Single play of the game (123 participants; with 65 of these participating at the retention test after 8 weeks), Repeated play of the game (93 participants, 57 at retention), or the control Video condition (119 participants, with 77 at retention). All participants first completed the external bias measures. Participants in the Single play condition then played the game before taking the post-test bias measures. Participants in the Repeat play condition played the game once and returned to the lab 7-10 days later to play the game again before taking the external post-test bias. Participants in the Video condition watched the instructional Video and then took the post-test bias measures. All participants were contacted by email 8 weeks after their (final) training session, and used a link to remotely access the Qualtrics survey for the retention test.

RESULTS

Immediate Bias Recognition and Knowledge (Pre-test and Post-test)

Bias knowledge and recognition showed a change from training (See Figure 2), with learning demonstrated through a main effect of pre-post ($F(1,327)=741.23$, $p<.00005$, MSE=1.40). While there was no main effect of condition ($F(2,327)<1$, MSE=3.10), some evidence of differential effectiveness of the forms of training was seen in a marginally non-significant interaction between pre-post and condition ($F(4,327)=2.66$, $p=.07$). Planned comparisons showed significant increased knowledge bias for a Single play of the game ($t(120)=-16.43$, $p<.0005$, d=1.49), Repeat play of the game ($t(91)=-15.31$, $p<.0005$, d=1.60), and from the Video control condition ($t(116)=-15.30$, $p<.0005$, d=1.41). The effect sizes suggest substantial learning about the biases occurred in all three training conditions. Examining the post-test performance using the planned linear contrasts showed a difference between game conditions ($t(330)=2.28$, $p<.05$), and a significant training advantage for the combined game conditions versus the Video condition ($t(330)=6.99$, $p<.0005$) and a significant benefit for the Repeat play condition over the control Video ($t(332)=4.66$, $p<.0005$). Thus when measured at acquisition, there was a slight but statistically significant advantage for the Repeat game play condition.

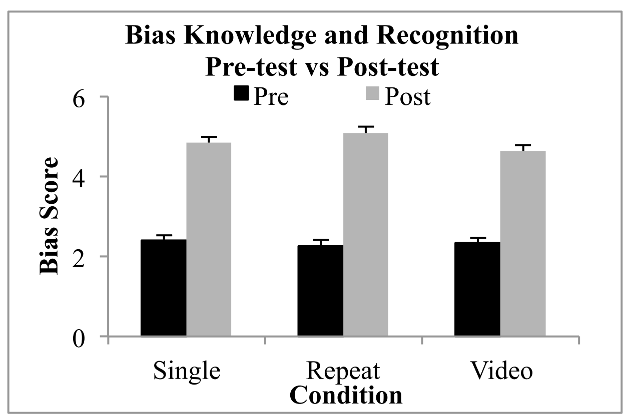

Figure 2. Bias Knowledge and Recognition by Condition for Pre-test and Post-test. Error bars show standard error.

Bias Recognition and Knowledge Retention (8 week post)

Bias knowledge and recognition showed evidence of greater change from the game conditions than the Video condition when assessed at the retention test (See Figure 3). Supporting the idea that at least some of the learning seen at the immediate post-test had been retained, there with a main effect of pre-test versus retention ($F(1,192)=70.30$, $p<.0005$, MSE=2.07). There was no main effect of condition ($F(2,192)=5.00$, $p<.01$, MSE=3.37), but differences between the conditions in the amount of learning retains was suggested by an interaction between pre-retention and condition ($F(2,192)=10.64$, $p<.0005$). Planned comparisons showed significant increase in knowledge of bias for the Single play condition ($t(63)=-3.31$, $p<.005$, d=0.41), Repeated play condition ($t(55)=-8.70$, $p<.0005$, d=1.16), and the Video control condition ($t(74)=-2.65$, $p<.05$, d=.31). These show a large effect size for learning about the biases from the Repeat play condition, and small to medium effect sizes from Single game play and Video conditions. Examining the retention test performance using the planned linear contrasts showed a difference between game conditions ($t(196)=-3.74$, $p<.0005$), a significant difference between combined game conditions versus the Video condition ($t(196)=3.17$, $p<.005$), and a significant difference between the Repeated play condition, and the Video condition ($t(196)=4.81$, $p<.0005$). As can be seen from these analyses greater retention of knowledge was present from the Repeated play of the game.

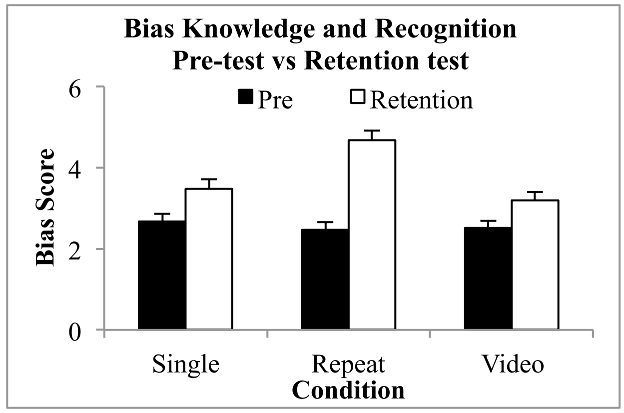

Figure 3. Bias Knowledge and Recognition by Condition for Pre-test and Retention test. Error bars show standard error.

Bias Mitigation

A DM MANOVA looking at all pre-test and immediate post-test with the three types of bias measures as dependent variables showed that some level of bias mitigation was present as shown by a significant pre-post training effect (Wilks' $\lambda=.48$, $F(3,328)=117.06$, $p<.0005$). Differences in the extent of this mitigation between the conditions was seen in a significant pre-post by condition (Single play game, Repeat play game, control Video) interaction (Wilks' $\lambda=.77$, $F(6,658)=15.51$, $p<.0005$). The same analysis for retention showed a significant pre-retention training effect (Wilks' $\lambda=.41$, $F(3,193)=45.10$, $p<.0005$), and a significant pre-retention by condition (Single play game, Repeat play game, control Video) interaction (Wilks' $\lambda=.88$, $F(6,386)=4.23$, $p<.0005$). These analyses suggest large overall training effects in bias mitigation, and differences in the levels of mitigation attained from the different forms of training, both immediately following acquisition and at the retention test. Follow-up analyses examined each bias measure individually.

Fundamental Attribution Error – Immediate Post-Test

FAE showed evidence of greater change in bias score from the game conditions than the Video condition (See Figure 4), with mitigation shown in a main effect of pre-post ($F(1,330)=305.12$, $p<.0005$, MSE=744), a main effect of condition ($F(1,330)=5.19$, $p<.01$, MSE=1593), and an interaction between pre-post and condition ($F(2,330)=17.28$, $p<.0005$) suggested that FAE mitigation differed across the conditions. Planned comparisons showed significant reductions in bias for the Single play condition ($t(122)=13.37$, $p<.0005$, d=1.21), Repeated play condition ($t(92)=8.27$, $p<.0005$, d=0.86), and the Video control condition ($t(118)=8.79$, $p<.0005$, d=0.81). Effect sizes suggest large immediate training effects for all three conditions. Examining the post-test performance using the planned linear contrasts showed a difference between game conditions ($t(332)=-2.54$, $p<.05$), a significant difference between combined game conditions versus the Video condition ($t(332)=-4.14$, $p<.0005$), and a significant difference between the Repeat play versus the Video condition ($t(332)=-2.17$, $p<.05$).

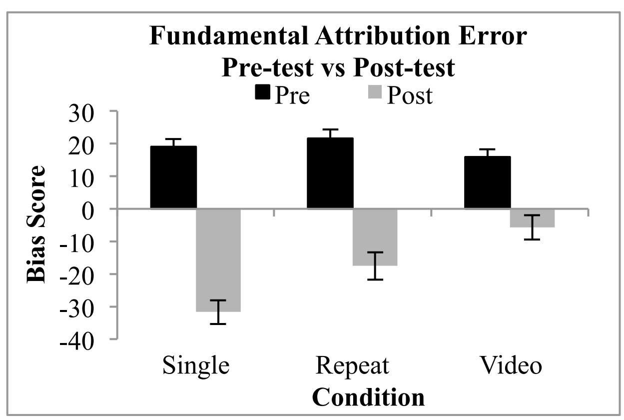

Figure 4. Fundamental Attribution Error by Condition for Pre-test and Post-test. Error bars show standard error.

The Single play of the game induced a greater negative shift in bias scores than Repeated play, and both had a larger magnitude effect on FAE than the Video.

Fundamental Attribution Error - Retention

FAE showed evidence of equally effective change from the game conditions and the Video condition (See Figure 5), with a main effect of pre-test versus retention ($F(1,195)=83.89$, $p<.0005$, MSE=934), no main effect of condition ($F(2,195) <1$, MSE=1467), and no interaction between pre-retention and condition ($F(2,195)=1.97$, $p = .14$). Planned comparisons showed significant reductions in bias for the Single play condition ($t(64)=5.03$, $p<.0005$, d=.62), Repeated play condition ($t(56)=5.76$, $p<.0005$, d=.76), and the Video control condition ($t(76)=4.68$, $p<.0005$, d=.53). The effect sizes are consistent with good retention of all three forms of training. Examining the retention test performance using the planned linear contrasts showed no difference between game conditions ($t(196)=1.02$, $p >.05$), no significant difference between combined game conditions versus the Video condition ($t(196)=-0.82$ $p>.05$), and no significant difference between the Repeated play condition and the Video condition ($t(196)=-1.21$, $p>.05$). Overall the FAE data reflect substantial retention in all conditions. Potential ceiling effects may limit the ability to detect differential performance across groups.

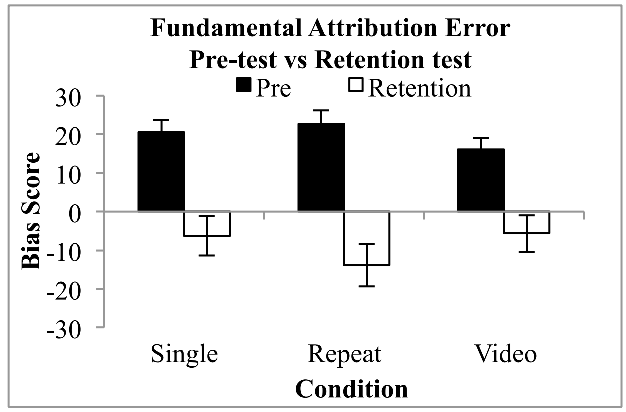

Figure 5. Fundamental Attribution Error by Condition for Pre-test and Retention test. Error bars show standard error.

Confirmation Bias – Immediate Post-Test

CB showed evidence of greater change in bias from the game conditions than the Video condition (See Figure 6), with a main effect of pre-post ($F(1,332)=120.58$, $p<.0005$, MSE=181), a main effect of condition ($F(2,332) =3.63$, $p<.05$, MSE=406), and an interaction between pre-post and condition ($F(2,332)=37.59$, $p<.0005$). Planned comparisons showed significant reductions in bias for the Single play condition ($t(122)=9.14$, $p<.0005$, d=.82), and the Repeated play condition ($t(92)=8.26$, $p<.0005$, d=.86), but no evidence of learning in the Video control condition ($t(118)=-0.80$, $p >.05$, d=-.07). The effect sizes show substantial reductions in confirmation bias from playing the game, but no effect from the training Video. Examining the post-test performance using the planned linear contrasts showed no difference between game conditions ($t(332)=0.78$, $p >.05$), but a significant training advance for the combined game conditions versus the Video condition ($t(332)=-7.08$, $p<.0005$) and a significant benefit for the Repeat play condition over the control Video ($t(332)=-6.24$, $p<.0005$).

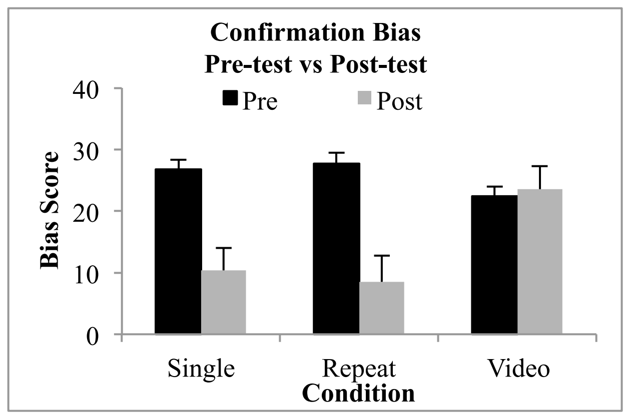

Figure 6. Confirmation Bias by Condition for Pre-test and Post-test. Error bars show standard error.

Confirmation Bias - Retention

CB showed evidence of greater change from the best game condition than the Video condition (See Figure 7), with a main effect of pre-test to retention test ($F(1,195)=42.57$, $p<.0005$, MSE=253), a main effect of condition ($F(2,195) =4.38$, $p<.05$, MSE=457), and an interaction between pre-retention and condition ($F(2,195)=11.12$, $p<.0005$). Planned comparisons showed significant reductions in bias for the Single play condition ($t(64)=2.48$, $p<.05$, d=.31), and the Repeated play condition ($t(56)=6.80$, $p<.0005$, d=.90), but no evidence of learning in the Video control condition ($t(76)=1.33$, $p>.05$, d=0.15). Effect sizes show large mitigation of confirmation bias from the Repeated play, a small to medium effect from Single game play, and a small non-significant mitigation effect from the Video condition. Examining the retention test performance using the planned linear contrasts showed a difference between game conditions ($t(196)=4.09$, $p<.0005$), no overall training advantage for the combined game conditions versus the Video condition ($t(196)=-1.58$, $p>.05$), but a significant difference between the best game condition, the Repeated play condition, and the Video condition ($t(196)=-3.44$ $p<.005$). Overall the CB data reflect the benefit of Repeat play for retention.

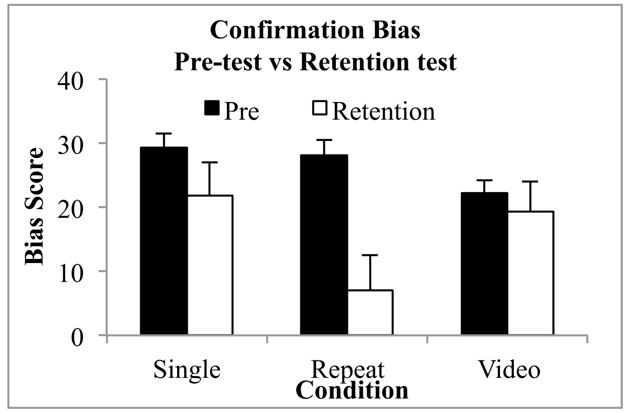

Figure 7. Confirmation Bias by Condition for Pre-test and Retention test. Error bars show standard error.

Bias Blind Spot – Immediate Post-Test

BBS showed evidence of greater change from the game conditions than the Video condition for the omnibus analyses (See Figure 8). There was a significant main effect of pre-post ($F(1,330)=12.96$, $p<.0005$, MSE=73), a main effect of condition ($F(2,330)=5.01$, $p<.01$, MSE=176), and an interaction between pre-post and condition ($F(2,330)=6.53$, $p<.005$). Planned comparisons showed significant reductions in bias for the Single play condition ($t(122)=3.51$, $p<.005$, d=.32), and the Repeated play condition ($t(92)=3.10$, $p<.005$, d=.32), but no evidence of learning in the Video control condition ($t(116)=-0.95$, $p>.05$, d=-0.01). Effect sizes reflect small to medium changes in bias blind spot following game play, but no evidence of mitigation from the Video. Examining the post-test performance using the planned linear contrasts showed no difference between game conditions ($t(332)=-1.30$, $p>.05$), but a significant training advance for the combined game conditions versus the Video condition ($t(332)=-4.54$, $p<.0005$) and a significant benefit for the Repeat play condition over the control Video ($t(332)=-3.12$, $p<.005$).

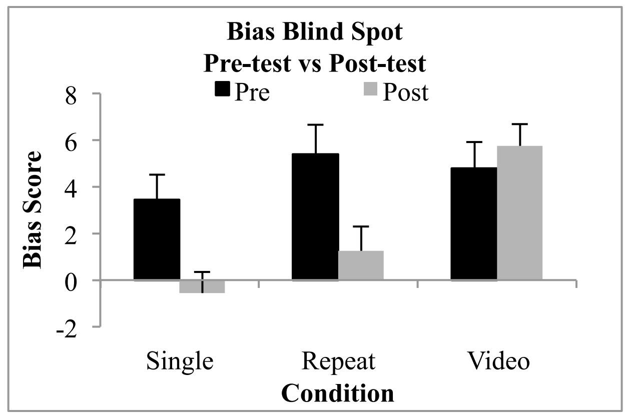

Figure 8. Bias Blind Spot by Condition for Pre-test and Post-test. Error bars show standard error.

Bias Blind Spot - Retention

BBS showed no evidence of greater change from the game conditions than the Video condition (See Figure 9). There was, with a main effect of pre-test versus retention ($F(1,195)=16.20$, $p<.0005$, MSE=92), no main effect of condition ($F(2,195)=2.06$, $p=.13$, MSE=164), and no interaction between pre-retention and condition ($F(2,195)=<1$). Planned comparisons showed significant reductions in bias for the Single play condition ($t(64)=2.71$, $p<.01$, d=.34), Repeated play condition ($t(56)=2.14$, $p<.05$, d=.28), and the Video control condition ($t(75)=2.06$, $p<.05$, d=.24). Effect sizes show small to medium mitigation of BBS at retention. Interestingly, a training effect is now seen also for the Video condition that was not observed at acquisition. This may be associated with the contradictory nature of training people on bias mitigation while presenting them a message about bias blind spot that suggested they were still biased. Examining the retention test performance using the planned linear contrasts showed no difference between game conditions ($t(196)=-1.14$, $p>.05$), no significant difference between combined game conditions versus the Video condition ($t(196)=-1.75$ $p=.08$), and no significant difference between the best game condition, the Repeated play condition, and the Video condition ($t(196)=-0.87$ $p>.05$).

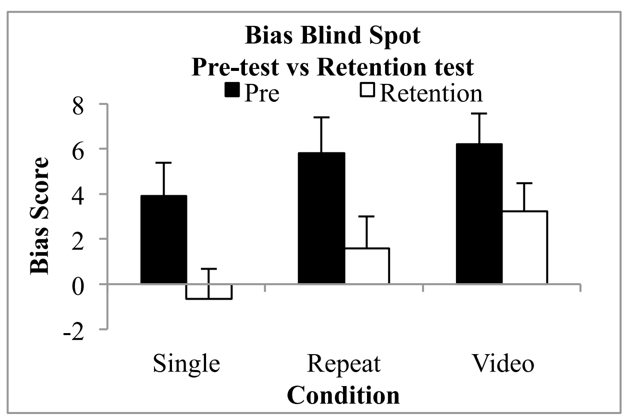

Figure 9. Bias Blind Spot by Condition for Pre-test and Retention test. Error bars show standard error.

Independent Assessment

An additional assessment of the effectiveness of the game was conducted by an independent lab (Bush, 2014). These tests employed novel measures of cognitive bias, unknown to the research team who developed the game, that featured different matched materials at pre-test, post-test, and retention test. All participants were recruited and run by the independent lab. Participants in the independent assessment engaged in the Repeat play condition (n=53) and the Video control condition (n=58), and also completed a similar 8-week retention test. Game play resulted in greater improvements on the bias measures than the control Video on all three biases at retention and for bias blind spot on the immediate post-test as well (FAE: Post-test 24% improvement vs. Video 19%, Retention test 15% vs. -3%, t-test reported $p<.05$; CB: Post-test 24% improvement vs. Video 14%, Retention test 33% vs. 16%, t-test reported $p<.01$; BBS: Post-test 17% vs. Video 4%, t-test reported $p<.05$, Retention test 27% vs. 11%, t-test reported $p<.01$).

DISCUSSION

Overall these results suggest highly promising performance from the bias training game developed for this project. The game proved an effective training method to reduce bias in a durable fashion. Reductions in bias were shown on the external questionnaire items, both immediately following training and when retested 8 weeks subsequent to training. In comparison to the professionally developed video, the game shifted decision making further from the tendency to affirm dispositional explanations in fundamental attribution error on the immediate post-test. The game reduced confirmation bias on both post-test and retention test whereas the video had no effect. Furthermore the game reduced bias blind spot to a greater extent on the immediate post-test than the video. These results broadly align with an independent assessment of the game's effectiveness that also found changes in all three forms of bias. Such changes from relatively brief (~40 minutes) training that transfer outside of the learning environment to items not just in a different context but of a different nature, support the notion that serious computer games can serve as effective training tools. In addition the results suggest that this current game has potential value in training bias mitigation within a wide spectrum of activities and jobs that that can be seen to be vulnerable to such biases. The ubiquity of confirmation bias and fundamental attribution error means there are a wide range of situations in which individuals could benefit from the type of rapid, generalizable bias mitigation that this training appears to offer.

The effect of repetition of the game was hypothesized to benefit retention, but this was only apparent within confirmation bias. However, the large retention effects observed in the single play condition for both fundamental attribution error and bias blind spot may have produced ceiling effects that limited the ability to detect differences between conditions.

This study demonstrates the effectiveness of using a game to train higher order cognitive skills such as biases and decision-making in particular. The design might also negate some other potentially valuable aspects of games as training platforms, given that participants' learning was precisely proscribed within the experiment. As suggested by current scholarship, games' tendency to generate intrinsic motivation may contribute to improved learning outcomes (Deci & Ryan, 1985; Salen & Zimmerman, 2003), and these same properties may also critically influence the decision to engage or re-engage in the training where the decision to attempt to learn rests with the learner. Moreover, the interactive characteristics of games that include failure and replay, along with the potential to explore different activities, paths, and outcomes, may contribute to learning, but also encourage participants to conduct refresher training. Research suggests that enabling learners to have greater control over their learning experience improves knowledge and increases motivation to learn (Reigeluth & Stein, 1983; Williams, 1996), which again can be features of game-based training. The importance to training outcomes of a choice to participate in additional training was apparent in the study of web-based refresher training for CPR skills in which Compton, Nafziger, Brown, Ottingham, and Moore (2005) found that a mere 3% of participants even accessed the website. The extent to which games can promote and encourage engagement in training remains an important issue for future research.

ACKNOWLEDGEMENTS

This work was supported by the Intelligence Advanced Research Projects Activity (IARPA) via the Air Force Research Laboratory contract number FA8650-11-C-7176 (PI: Tomek Strzalkowski). 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.

REFERENCES

Annett, J. (1979). Memory for skill. In M. M. Gruneberg & P. E. Morris (Eds.) *Applied problems in memory* (pp. 215-247). London: Academic Press.

Arnott, D. (2006). Cognitive biases and decision support systems development: a design science approach. *Information Systems Journal*, *16*(1), 55-78.

Boyd, J. R. (1995). The Essence of Winning and Losing, 28 June 1995 a five slide set by Boyd. Retrieved from http://www.danford.net/boyd/essence.htm.

Bush, R. M. (2014). Tuning the knobs and dials: Empirically maximizing features for serious games. *Games+Learning+Society 10 Conference.* Madison, WI, June 10th-13th 2014.

Cepeda, N, Vul, E, Rohrer, D, Wixted, J, & Pashler, H. (2008). Spacing effect in learning: A temporal ridgeline of optimal retention. P*sychological Science, 19,* 1095-1102.

Compton, S. Nafziger, S.D., Brown, L.H., Ottingham, L.V., & Moore, R. (2005). Internet-based cardiopulmonary resuscitation/automated external defibrillator refresher training. *Academic Emergency Medicine, 12(5)*, 73a.

Cook, M. B. & Smallman, H. S. (2008). Human factors of the confirmation bias in intelligence analysis: Decision support from graphical evidence landscapes. *Human Factors, 50,* 745-54.

Deci, E, & Ryan, R. (1985). *Intrinsic motivation and self-determination in human behavior*. New York: Springer.

Fischhoff, B. (1982). Debiasing. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge: Cambridge University Press: 422 – 444

Gigerenzer, G., & Brighton, H. (2009). Homo heuristicus: Why biased minds make better inferences. *Topics in Cognitive Science*, 1, 107-143.

Gilbert, D. T. & Malone, P. S. (1995). The correspondence bias. *Psychological Bulletin 117 (1),* 21–38.

Holding, D.H. (1991). Transfer of training. In J. E. Morrison (Ed.), *Training for Performance: Principles of Applied Human Learning* (pp. 93-125). NY, NY: John Wiley

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

Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist*, 58, 697– 720.

Kahneman, D. (2011). *Thinking, Fast and Slow*. New York: Macmillan.

Kahneman, D., Lovallo, D., Sibony, O. (2011). Before you make that big decision. *Harvard Business Review*, 89(6), 50-60.

Klayman, J. & Brown, K. (1993). Debias the environment instead of the judge: An alternative approach to improving diagnostic (and other) judgment. *Cognition*, 49, 97-122.

Larrick, R. P. (2004). Debiasing. In D. J. Koehler & N. Harvey (Eds.), *Blackwell handbook of judgment and decision making*. (pp. 316-338). Malden, MA: Blackwell Publishing.

Mezirow, J. (1991). *Transformative dimensions of adult learning*. San Francisco, CA: Jossey- Bass.

Nickerson, R, S. (1998). Confirmation Bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, 2(2), 175–220.

Pronin, E., Lin, D. Y., & Ross, L. (2002). The bias blind spot: Perceptions of bias in self versus others. *Personal Social Psychology Bulletin,* 28(3), 369-381.

Pronin, E. (2007). Perception and misperception of bias in human judgment. *Trends in Cognitive Sciences*, 11, 37-43.

Reigeluth, C.M., & Stein, F. S. (1983). The elaboration theory of instruction. In C.M. Reigeluth, ed., *Instructional design theories and models* (pp. 335-82). Hillsdale, NJ: Erlbaum.

Reyna, V. F., Chick, C. F., Corbin, J. C., & Hsia, A. N. (2014). Developmental reversals in risky decision-making: Intelligence agents show larger decision biases than college students. *Psychological Science, 25(1),* 76-84.

Roediger, H. L., & Karpicke, J. D. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science*, *1*(3), 181-210

Salen, N, & Zimmerman, E. (2003). *Rules of play.* Cambridge, MA: MIT Press.

Sanna, L.J., Schwarz, N., & Stocker, S.L. (2002). When debiasing backfires: Accessible content and accessibility experiences in debiasing hindsight. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 497–502.

Schmidt, R.A. & Bjork, R.A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science*, 3, 207-217.

Slamecka, N. J. & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory, 4(6),* 592-604.

Tetlock, P.E. (1985). Accountability: A social check on the fundamental attribution error. *Social Psychology Quarterly*, 48(3), 227-236.

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

West, R. F., Meserve, R. J., & Stanovich, K. E. (2012). Cognitive sophistication does not attenuate the bias blind spot, *Journal of Personality and Social Psychology*, 103(3), 506-519.