Title

Implicit and explicit training in the mitigation of cognitive bias through the use of a serious game

Authors

Dunbar, Norah E
Miller, Claude H
Adame, Bradley J
et al.

Publication Date

2014-08-01

DOI

10.1016/j.chb.2014.04.053

Peer reviewed

Computers in Human Behavior 37 (2014) 307–318

Contents lists available at ScienceDirect

Computers in Human Behavior

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

Implicit and explicit training in the mitigation of cognitive bias through the use of a serious game

Norah E. Dunbar [a,1,*], Claude H. Miller [a], Bradley J. Adame [d], Javier Elizondo [a], Scott N. Wilson [a], Brianna L. Lane [a], Abigail Allums Kauffman [e], Elena Bessarabova [a], Matthew L. Jensen [a], Sara K. Straub [a], Yu-Hao Lee [f], Judee K. Burgoon [b], Joseph J. Valacich [b], Jeffrey Jenkins [g], Jun Zhang [c]

[a] *University of Oklahoma, United States*
[b] *University of Arizona, United States*
[c] *University of Michigan, United States*
[d] *Arizona State University, Hugh Downs School of Human Communication Stauffer Hall Building A, Room 412 P.O. Box 871205, Tempe, AZ 85287, USA*
[e] *The University of Texas of the Permian Basin Development Office, MB 4230A, 4901 E. University, Odessa, Texas 79762, USA*
[f] *University of Florida, Department of Telecommunication, P.O. Box 118400, Gainesville, FL 32611-8400, USA*
[g] *Brigham Young University, Marriott School of Management, Department of Information Systems (790 TNRB) Provo, UT 84602-3113, USA*

ARTICLE INFO

Article history:
Available online 14 June 2014

Keywords:
Video games
Cognitive biases
Training
Bias mitigation
Instructional testing

ABSTRACT

Heuristics can interfere with information processing and hinder decision-making when more systematic processes that might lead to better decisions are ignored. Based on the heuristic-systematic model (HSM) of information processing, a serious training game (called *MACBETH*) was designed to address and mitigate cognitive biases that interfere with the analysis of evidence and the generation of hypotheses. Two biases are the focus of this paper—*fundamental attribution error* and *confirmation bias*. The efficacy of the serious game on knowledge and mitigation of biases was examined using an experiment in which participants ($N = 703$) either played the *MACBETH* game or watched an instructional video about the biases. Results demonstrate the game to be more effective than the video at mitigating cognitive biases when explicit training methods are combined with repetitive play. Moreover, explicit instruction within the game provided greater familiarity and knowledge of the biases relative to implicit instruction. Suggestions for game development for purposes of enhancing cognitive processing and bias mitigation based on the *MACBETH* game design are discussed.

1. Introduction

Deliberative decision-making requires time and cognitive effort; therefore people regularly rely on heuristics—mental shortcuts—to make fast decisions. Heuristic processing of information may not be a problem when the stakes are low (i.e., when the cost of being wrong is insignificant) or when the heuristic aligns with the context. However, there are situations in which careful reasoning is required to make informed decisions, and over-reliance or misapplication of heuristics can lead to systematic cognitive biases and catastrophic outcomes (e.g., poor medical treatment, bad public policy, or even threatened national security).

Cognitive biases, or distortions in patterns of thinking, are very difficult to mitigate because people are usually unaware of their presence and operation. Moreover, decision makers are especially prone to biases when making relatively uncertain evaluations requiring large amounts of cognitive effort (Abelson & Levi, 1985). Research has shown that professionals and experts are no less likely to commit cognitive biases than ordinary people when making important decisions (Englich, Mussweiler, & Strack, 2006). Given the propensity for cognitive biases to short-circuit the effectiveness of everyday decision-making, the need for methods to mitigate their effects is constant. To this end, the present study examines the use of a serious video game to train people, such as intelligence analysts and policy makers, in the mitigation of cognitive biases. Intelligence analysts in particular must make quick decisions with very little information, and so are especially susceptible to cognitive biases (Heuer, 1999).

To increase systematic decision-making and reduce the prevalence of cognitive biases when analyzing intelligence information,

* Corresponding author. Address: Department of Communication, University of Oklahoma, Burton Hall 101, Norman, OK 73019, United States. Tel.: +1 805 893 4517.
E-mail address: ndunbar@comm.ucsb.edu (N.E. Dunbar).
[1] Present address: University of California Santa Barbara, Dept. of Communication, Social Sciences & Media Studies Bldg (SSMS), Santa Barbara, CA 93106-4020.

http://dx.doi.org/10.1016/j.chb.2014.04.053
0747-5632/

we created a serious video game called *MACBETH* (Mitigating Analyst Cognitive Bias by Eliminating Task Heuristics). This paper presents the results of a multi-site experiment testing the ability of *MACBETH* to mitigate two specific biases described in the next section—the *fundamental attribution error* (FAE) and *confirmation bias* (CB).[2] Although others have attempted to reduce bias through a variety of training methods, the effectiveness of those attempts has been minimal and their lasting effects remain unclear (Dunbar et al., 2013). To our knowledge, no other experimental studies of video game training for the mitigation of cognitive bias have been published. The *MACBETH* case study can provide guidance for researchers attempting to train professionals about the nature of cognitive bias, while providing evidence for the efficacy of a novel, serious video game-based training method.

2. Theoretical approach

A primary causal mechanism cited for biased information processing is the reliance on heuristic social information processing, which, as described by Chaiken's heuristic-systematic model (HSM; Chaiken, 1980; Todorov, Chaiken, & Henderson, 2002), is a nonanalytic orientation relying on a quick and minimally careful consideration of informational cues. The HSM defines heuristics as mental shortcuts, or simple decision rules arising from conventional beliefs and expectations used repeatedly in daily interactions. In contrast to heuristic processing, systematic social information processing requires more careful consideration of all available evidence and is much more cognitively taxing (Chen & Chaiken, 1999). An over-reliance on heuristics when useful information is available can lead to biased information-processing and a range of suboptimal decisional outcomes, including faulty reasoning, and a failure to make sound credibility judgments. People may often erroneously believe they are making decisions based on sound evidence when in fact they are actually making guesses. According to the HSM, only if adequately motivated, with sufficient time and ability to process information, will individuals choose to engage in systematic processing. Thus, both the motivation and ability to process information are critical for reducing analytical over-reliance on simple heuristics (Todorov et al., 2002).

The *MACBETH* game was designed to train players on two particular biases: FAE and CB, the later of which is the tendency to search for or interpret information in a way that confirms one's preconceived assumptions, biases, expectations, or hypotheses (Nickerson, 1998). When faced with multiple possibilities, CB lowers the probability that one's initial hypothesis will be rejected (Oswald & Grosjean, 2004; Watson, 1960).

Several studies have examined ways to mitigate the negative effects of CB in similar investigation tasks (Hill, Memon, & McGeorge, 2008; Krems & Zierer, 1994; Oswald & Grosjean, 2004; Rassin, 2010). For example, O'Brien's (2009) study of CB in criminal investigations demonstrated that participants who considered why their hypothesis might be wrong showed less bias, whereas those who generated several additional hypotheses did not, suggesting that too many alternatives or too much high task complexity can hinder one's systematic processing ability. In more complex tasks such as the intelligence gathering and analysis task tested here, information consistent with alternative hypotheses has the effect of prematurely reifying the original hypothesis (O'Brien, 2009). Games are ideal settings for training about CB because players can be prompted to offer alternative hypotheses when new information is received or can be encouraged to delay making hypotheses before sufficient information is known.

[2] The game was also designed to mitigate the bias blind spot (BBS), but the BBS results are presented elsewhere due to space limitations and because a different game mechanic not described here was used for BBS.

The second bias we examined was the FAE, which is the tendency for people to over-emphasize stable, personality-based explanations for behaviors observed in others—referred to as dispositions—while under-emphasizing the role and power of transitory, situational influences on the same behavior (Harvey, Town, & Yarkin, 1981; Mowday, 1981). However, if a behavior is truly caused more by dispositions or other personality variables, then using those dispositions to explain a behavior would not be in error (Gifford & Hine, 1997). The problem lies in the general tendency of humans to overlook situational variables while emphasizing and prematurely attributing causes to dispositions, regardless of whether they are the true or only causes of the behavior in question. This cognitive neglect can be explained by the HSM, since dispositional attributions are simpler, demand less effort, and may satisfy decisional needs more easily than expending the time and energy required to investigate situational and/or contextual explanations for relevant behaviors. For example, if you discover a falsehood, it is easier to explain that behavior by calling someone a "liar" than to uncover the reasons behind the lie (O'Sullivan, 2003). This form of FAE is exacerbated when making judgments about the motives of others, since it is more difficult to access the situational factors associated with others' behaviors relative to one's own behaviors.

As with the mitigation strategies for CB, exploring other situational hypotheses should make analysts aware of their tendency to rely on dispositions, and instead, make them rely more on well-reasoned judgments. In addition, Hodgins and Knee (2002) suggested that openness to experience, an aspect of mindfulness (characterized by attentiveness and awareness), attenuates cognitive defensiveness such as using self-serving bias and stereotyping. Another study found that mindfulness when receiving a negative evaluation can reduce hostile attribution bias and aggressiveness (Heppner et al., 2008). Games can train players to learn the difference between situational and dispositional information and mitigate the FAE by encouraging mindfulness in decision-making through rewards in the game.

3. Experiential learning of biases through video games

Video games are ideal media for learning and mitigating cognitive biases because their interactivity facilitates experiential learning. Hands-on experience is at the heart of many learning theories (see Kolb, 1984). These theories argue that learners will have a deeper understanding of the issue through the experience of problem solving, experimenting with different solutions, and observing the consequences of their decisions. According to Hoover and Whitehead (1975) "Experiential learning exists when a personally responsible participant cognitively, affectively, and behaviorally processes knowledge, skills, and/or attitudes in a learning situation characterized by a high level of active involvement" (p. 25). Since decision-makers are often unaware of their cognitive biases, experiential learning is necessary for mitigating cognitive biases because it allows decision-makers to actively process information, make decisions, and observe the consequences of their actions. Since decision makers are personally responsible for their decisions and resulting consequences, this experience is more likely to make them aware of their own biases and actively practice mitigating them. This is especially true in the area of intelligence analysis where mistakes can be grave, so making and learning from these mistakes in a safe environment is more desirable than learning about them from suffering their consequences in the real world.

Video games can support experiential learning through simulating real decision-making scenarios, providing dynamic feedback, providing opportunities to experiment with different action, and allowing players to observe the consequences (Jarmon, Traphagan, Mayrath, & Trivedi, 2009). In comparison to traditional training tools which are more passive, such as an instructional video or a lecture, video games are more interactive because they require active decision-making from players. We hypothesize that an interactive serious game will be more effective in mitigating cognitive bias than a traditional, passive, instructional video training because it engages the players in experiential learning.

H1. A serious video game will be superior to an instructional video at mitigating CB and FAE.

4. Implicit vs. explicit training in serious games

Although video games are often considered entertainment, and have long been eschewed by many serious educators, they have been demonstrated to enhance and advance traditional models of learning, even for complex tasks such as the mitigation of cognitive biases (Ciavarro, Dobson, & Goodman, 2008; Day, Arthur, & Gettman, 2001; Masson, Bub, & Lalonde, 2011; Squire, 2003). Video games are generally more engaging than traditional modes of learning because games are better suited to facilitate more dynamic, active, *implicit* learning modes, whereas traditional forms of instruction—such as classroom lectures, instructional videos, or practice lessons—are more suited for relatively static, passive, *explicit* learning (Gee, 2011; Squire, 2003). Implicit learning refers to the unintentional acquisition of complex knowledge or skills, often through problem-solving or experiential learning. In contrast, explicit learning involves intentional acquisition of information in the form of declarative knowledge, often through memorization of knowledge (Ciavarro et al., 2008; Raab, 2003).

Although some researchers have argued that games with implicit learning provide a more enjoyable experience by stimulating intrinsic motivation for the learning task (Ciavarro et al., 2008; Tüzün, Yılmaz-Soylu, Karakuş, İnal, & Kızılkaya, 2009), the question of the effectiveness of the game is paramount. Based on their review of the literature on implicit or explicit learning, Habgood, Ainsworth, and Benford (2005) recommended that game designers should strive to achieve two goals: (1) avoid interrupting the flow of the game by delivering learning material through the game segments that are the most fun to play, and (2) embody the learning material within the structure of the game by providing an external representation of the learning content that is explored through the core mechanics of game play. However, they acknowledged there is no definitive evidence suggesting such an approach will necessarily produce more effective learning, and indeed, there is a chance it might actually produce *less* effective learning if the immersion in the game inhibits that player's ability to think about the lessons being taught.

One potential difficulty in determining whether implicit vs. explicit training will be more effective in mitigating cognitive biases is the complexity of the task. Compared to relatively simple concepts tested in other educational games such as geographical locations of countries (Tüzün et al., 2009), or following the rules of hockey (Ciavarro et al., 2008), learning to reduce one's reliance on cognitive biases is highly complex and relatively difficult to achieve given the fact that biases are ingrained behaviors that are a natural and practical aspect of human cognitive processing (Kahneman, 2011). Across four experiments, using different sports-specific decision-making tasks, Raab (2003) found that implicit learning was better in low-complexity situations and explicit learning was better in high-complexity situations. In addition, the HSM predicts that when greater cognitive resources are needed, one is more likely to rely on heuristics, which may lead to biased decision making. Thus, given the complexity of the bias training task, by minimizing the cognitive resources needed to learn the lesson through explicit training, the better players will be at assimilating knowledge about the biases, and considering the mitigation strategies provided. We test this reasoning both by comparing a video game to a training video via H1, but also by testing a video game with explicit training to a video game with implicit training via the following hypothesis:

H2. Game play with explicit training increases (a) familiarity with biases, (b) bias knowledge, and (c) reduction in biased judgments compared to game play with implicit training.

Bias mitigation involves more than merely increased knowledge; rather, it is demonstrated by improved application of that knowledge resulting in unbiased judgments. Together, bias knowledge and performance (i.e., application of bias mitigating strategies) form the basis for improved decision-making competency. Given the complexity of the bias mitigation task, we propose that longer game play-time and repeated play should enhance the game's ability to improve knowledge of biases and reduce biased judgment. Thus, we predict:

H3. Longer exposure to a video game through (a) repeated play or (b) longer duration of play is more effective at mitigating CB and FAE.

5. Method

5.1. Participants

Participants included 703 students recruited by mass emails and classroom announcements at two different U.S. Universities, both with populations of roughly 30,000 students; one located in the south-central U.S. (called University 1; $n = 311$) and the other in the southwest (called University 2; $n = 392$). We included more than one location to determine the robustness of our findings and to ensure the results were not geographically specific. The sample included 47% females, and ranged from 18 to 62 years of age ($M = 22.03$, $SD = 5.34$). Participants self-reported as 59% white, 24% Asian, 8% Hispanic, 4% African American, 2%, Native American, and 2% other. English was the first language for 77% of all participants, who had completed an average of 3.24 years of education ($SD = 1.86$) beyond high school.

The two samples were drawn from somewhat different populations. The University 1 sample was recruited through a mass email sent to the entire University and the University 2 sample was drawn from students in Management Information Systems and the College of Management. It should also be noted that the University 2 sample had a larger proportion of participants for whom English was a second language (29.3%) relative to the University 1 sample (13.9%). Specifically, the University 2 sample had a larger number of students from India ($n = 58$) and China ($n = 39$) compared to University 1 ($n = 3$ and 10 respectively). As much as possible, research associates at both locations followed an identical experimental script and procedures, and location effects were tested in all analyses to determine whether the diversity of the sample moderated the effectiveness of the *MACBETH* game.

5.2. Procedure

Participants were notified of the opportunity to participate in the experiment through email and in-class recruitments, whereupon appointments were arranged using a third-party scheduling service. An IRB approved email message was sent directing partic-

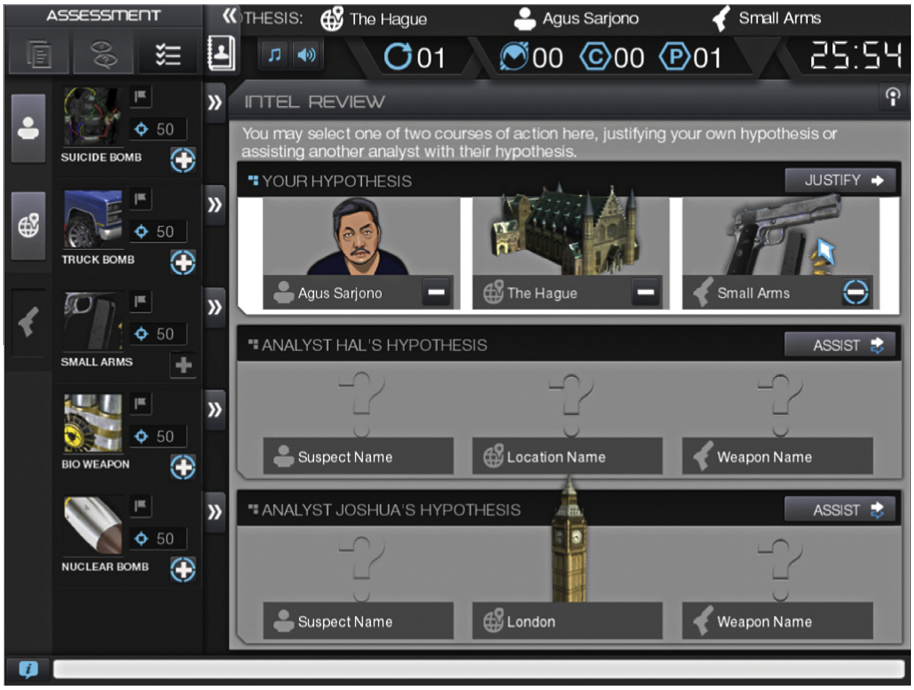

Fig. 1. The Intel Review screen of the game showing the player's suspect, location, and weapon hypothesis.

ipants to an on-line pre-survey assessing demographics and personality measures, after which participants were directed to the scheduling website where they were scheduled for lab sessions in groups of 20 participants at each location. A total of 958 participants completed the pre-survey, of whom 753 reported for their lab appointments, resulting in a retention rate of 73%.

Upon arriving at the lab, an experimenter greeted participants and randomly assigned them to one of the 10 experimental conditions in blocks to avoid participants in different duration conditions participating together. Participants were then directed to another room in the lab where experimenters administered pretest measures via an on-line survey, followed by the experimental treatment (either one of the *MACBETH* game conditions or the instructional video), after which post-treatment measures were administered via a second on-line survey. At the end of their first session, participants assigned to the repeat play condition scheduled their next appointment to be completed in the lab within one week; whereas those assigned to the take-home condition were given login instructions for accessing the game online. All participants were paid $20 for their participation in each lab session and were reminded they would be emailed a link to a follow-up survey in 8 weeks. Eight weeks from the date they last played the game, whether in the lab or at home, participants were emailed a link to a final posttest survey measuring the same outcomes using alternative scale item variations, upon completion of which they were emailed a $30 on-line gift card (University 1) or returned to the lab to receive a $30 cash payment (University 2).

5.3. Experimental conditions: MACBETH and the instructional video

In the *MACBETH* game, players, called "analysts," are presented with a fictional scenario of an impending terrorist attack, and their task is to figure out who the suspect is, where the attack will occur, and what method of attack will be used. *MACBETH* is a turn-based game, where a human participant plays cooperatively with two non-playable characters (NPCs). In any one turn, analysts are able to gather two pieces of information about the suspect, location, and/or weapon from a combination of intel sources (see Fig. 1). After gathering information, the human player can generate a hypothesis or aid another analyst (an NPC) if they have information proving or disproving the other analyst's current hypothesis. Throughout the game, analysts learn about the cognitive biases, and receive implicit and/or explicit feedback (based on condition) encouraging them to delay making a hypothesis, seek disconfirming information that can be used to disprove their hypothesis, and offer alternative hypotheses in their efforts to mitigate CB.

As part of the scenario, players are trained to mitigate FAE via the Archive minigame (a game inside the primary game—see Fig. 2). In Archive, players review past case files and make threat assessments on profiles of real-life individuals. Players are encouraged to rely more directly on situational, as opposed to dispositional, cues to mitigate FAE; correct answers in Archive are rewarded with resources that can be used to unlock additional intel. A more detailed development of the *MACBETH* game is beyond the scope of this paper, but can be found elsewhere (Dunbar et al., 2013).

In addition to examining bias mitigation via a serious videogame, one of the goals of this project was to test the effectiveness of the game compared to more traditional learning methods designed to mitigate cognitive biases. To this end, an instructional video was created by the funding agency[3] to act as a comparison condition for the experiment. The video was designed to inform viewers about the nature of cognitive biases by using entertaining vignettes that include a professorial host wearing a lab coat. Five "real-life" vignettes serve as examples of the protagonists revealing their cognitive biases. The host reviews each bias, along with possible mitigation strategies for each. Our experiment not only manipulates different versions of the game to determine the most effective game treatment, but also compares non-traditional active learning in an immersive game to the traditional, passive learning style associated with watching an instructional video.

[3] The Intelligence Advanced Research Projects Activity (IARPA) funded this work via their SIRIUS program through the Air Force Research Laboratory. They created the instructional video without input from the research team, who did not see the video until *MACBETH* development was nearly complete.

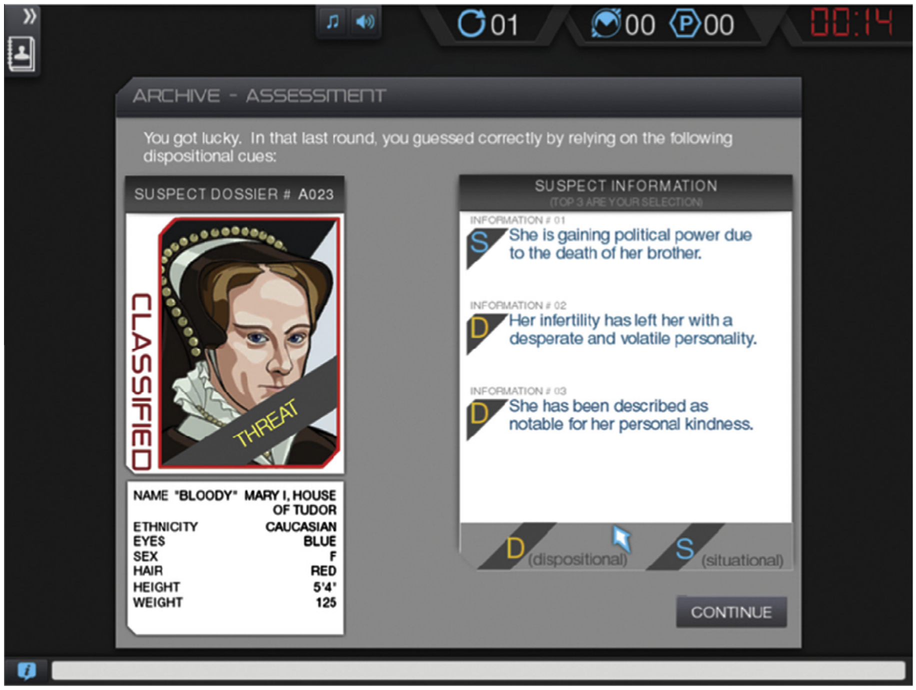

Fig. 2. The Archive minigame for teaching the difference between dispositional and situational cues.

5.4. Experimental manipulations of independent variables

Three different independent variables resulted in ten experimental conditions. The instructional video, created by IARPA, was tested against various treatments of the *MACBETH* game. The variables manipulated within the game included:

5.4.1. Implicit/explicit training

Participants played either a version of the game explicitly instructing them about the three biases, or, in the implicit condition, a game that did not overtly include this information. The explicit version of the game contained pop-up quizzes at various points in the game with text defining the biases, followed by multiple-choice questions checking whether they had learned the relevant definitions. If players did not answer the questions correctly, they were given the definition again along with a follow-up multiple-choice question. The FAE quiz appeared the first time players entered the Archive mini-game, and the CB quiz appeared the first time they entered the hypothesis-testing portion of the game. In the implicit condition, players simply began the game without receiving the bias definitions nor engaging in the multiple-choice questions.

5.4.2. Duration

Both 30- and 60-min versions of the game were created, and players were randomly assigned to one or the other condition. A clock counting down the time remaining in the game was presented in the center of the main menu screen at the outset of the game, and moved to the upper right corner of the screen during game play. It was visible at all times, except when the scenario menu was presented, and during the beginning and ending cinematics for each scenario. Players received a warning when they had 5 min remaining in their treatment. When the clock expired, they were instructed to submit their final hypothesis, thereby ending the game; after which they were shown a final outcome screen, and then informed that their treatment had ended.

5.4.3. Repetition

Players were randomly assigned to either a single-play (in-lab), or repeated-play condition. The repeated condition could be either in-lab or take-home. Repeat players were asked to sign up for a second lab visit one week following their first visit. Those in the repeat play condition who did not return for their second play (21%) were analyzed as single-play participants. Independent sample *t-tests* comparing the repeat-play return group to the repeat-play non-return group were conducted to ensure the two groups did not differ significantly on any of the pre-lab personality or demographic variables. No significant differences were found.

For the take-home condition, players were given login instructions requesting they play the game at least twice at home—or as often as they liked within the two weeks allowed (after which the game was disabled). Of the 102 players assigned to the take-home condition, 50 logged in at least one time from home and played an average of 53 min per session. Players averaged between 1 and 9 play sessions, and completed an average of 4 game scenarios each. To further assure that take-home players who played at home were no different from take-home players who did not play at home, we conducted independent sample *t* tests comparing the two groups on all pre- and posttest measures of bias knowledge and bias mitigation, and found no significant differences. Players assigned to the take-home condition who did not play at home were converted to and retained within the explicit, 60 min, single-play condition.

5.5. Measurement

As mentioned, demographic, personality and covariate measures were included in the pre-lab survey, and bias mitigation, familiarity, and knowledge questions were given in the pre- and posttest measures. To avoid testing effects on mitigation outcomes, all bias mitigation scale items were administered first, followed by familiarity and knowledge questions.

5.5.1. Personality variables and covariates

Given that the research literature suggests a variety of personality and demographic variables are related to the mitigation of cognitive bias, we measured a range of relevant variables for inclusion as covariates in our analyses.

5.5.1.1. Need for cognition. Research has shown that an aptitude for mathematics and/or logic (Stalder, 2000) and a high score on the decisiveness sub-factor of the need for cognition scale (NFC; Stalder, 2009) can attenuate the FAE. Likewise, a cultural tendency toward collectivism is known to reduce CB (Kastenmüller, Greitemeyer, Jonas, Fischer, & Frey, 2010).

5.5.1.2. Big 5 personality traits. We used Gosling et al.'s (2003) 10-item Personality Inventory (TIPI) to measure the "Big 5" personality traits. Although the five 2-item scales showed low internal consistency across traits, as expected, each pair of items was significantly correlated (openness [$r = .36$]; conscientiousness [$r = .39$]; extroversion [$r = .63$]; agreeableness [$r = .22$]; and emotional stability [$r = .48$]).

5.5.1.3. Cultural orientation. We used composite measures for the four dimensions, horizontal individualism (HI, 3-item $\alpha = .72$), vertical individualism (VI, 3-item $\alpha = .72$), horizontal collectivism (HC, 3-item $\alpha = .67$), and vertical collectivism (VC). The reliability for VC was below acceptable levels (3-item $\alpha = .47$), and thus eliminated from the analyses.

5.5.1.4. Need for closure. An 18-item subscale was adapted from a larger scale developed by Kruglanski, Webster and Klein (1993), measured on a 7-point Likert continuum anchored by strongly agree and strongly disagree ($\alpha = .82$).

5.5.1.5. Logic aptitude. We measured this construct using questions from published versions of the Law School Admission Test (LSAT) (Stalder, 2009). A significant portion of the LSAT is devoted to assessing one's ability to use logical reasoning in determining a valid conclusion. Participants were presented with instructions similar to those found in the LSAT, whereupon they answered three multiple-choice questions involving a passage containing various pieces of relevant and irrelevant information. Participants were presented with one of five logical conclusions and asked to select the one valid answer.

5.5.1.6. Computer comfort and gaming experience. We created scales to measure participants' comfort with both computers and gaming to determine if those variables moderate player experience with the *MACBETH* game. The two scales demonstrated good internal consistency (5-item $\alpha = .80$ for computer comfort and 5-item $\alpha = .91$ for gaming experience). Participants were also asked the types of games they play and the number of hours they play them per week. First-person shooter games, puzzle games, and sports games seemed to be the most common. On average, the sample spent 10.39 h per week playing games ($SD = 18.29$).

Because differences in genre type and total gaming hours between males and females are well documented in the literature (Hartmann & Klimmt, 2006), we tested for sex differences, which revealed a significant difference for sex on total gaming hours $t(409) = 2.60$, $p = .01$ (males $M = 12.52$, $SD = 22.93$; females $M = 7.8$, $SD = 9.74$), as well as on several of the genres. Males were more likely to play action, first-person shooter, strategy and sports games while females were more likely to play puzzle and rhythm games. Most relevant here is the difference on strategy games $t(427) = 3.77$, $p < .0001$ (males $M = 1.20$, $SD = 2.83$; females $M = 0.39$, $SD = 1.13$). Because *MACBETH* is considered to be a strategy game and males have more experience with this type of game, they may have a shorter learning curve. Thus, sex was included as a factor in all the following analyses.

Table 1
Cronbach's alpha reliabilities for bias mitigation scales.

Measure	Test period	α
NewCB	Pretest	.77
	Posttest1	.82
	Posttest2	.86
	8-week Posttest	.76
CBAM	Pretest	.69
	Posttest1	.74
	Posttest2	.88
	8-week posttest	.74
FAE Scenario Dispositional	Pretest	.70
	Posttest1	.86
	Posttest2	.87
	8-week Posttest	.83
FAE Scenario Situational	Pretest	.75
	Posttest1	.84
	Posttest2	.85
	8-week Posttest	.86
FAE Video Posttest 1	Contestant	.83
	Questioner	.75
FAE Video 8-week Posttest	Contestant	.80
	Questioner	.75

5.5.2. Dependent measures: bias familiarity, recognition, and knowledge

First, participants rated their degree of familiarity with each of the biases from very familiar to very unfamiliar on a 7-point scale. A second measure of bias recognition directed participants to match the definitions of CB and FAE along with four other biases with the appropriate bias names in a "drag and drop" exercise. It was scored such that participants earned one point for each correct match, so scores could range between 0 and 6 [4].

We also created a series of multiple choice exam-style questions to measure bias knowledge in which participants were given an example of a bias and directed to name it appropriately. For example: "Sarah is riding her bicycle through a busy part of town. Upon making a turn into a new street, she is cut off by the driver of a new luxury car. Sara immediately curses at the driver and assumes that the driver is irresponsible, uncaring, and a generally bad person." Participants received three of these questions for each bias at each test administration, for a possible score ranging between 0 and 6.

5.5.3. Dependent measures: bias mitigation measures

Previously published and validated scales for both CB and FAE were used as a starting point to determine appropriate measures of each bias. When only a few measures or single-item measures were available in the research literature, additional items were created so that the pre- and posttest measurement could be conducted using slightly differing item clones, which were pilot tested on a sample of undergraduate students at both universities ($N = 276$) to determine the comparability of the subsets of questions. Reliability of all scales was established prior to the experiment (see Table 1 for scale internal consistency across test periods).

[4] CB and FAE were presented along with four additional biases: bias blind spot (BBS), anchoring bias (AB), representativeness bias (RB), and projection bias (PB). Although only three biases were included in *MACBETH* (FAE, CB, and BBS), because knowing the definition of those biases would help find the definitions of the other three biases through process of elimination, we can still expect improvement after training on the overall scale.

5.5.3.1. CB. For CB, Rassin's (2010) test strategy scale (TSS) measure of confirmation proneness was used as a starting point. Because the original 10-item scale employing descriptions of situations failed to achieve satisfactory reliability in a measures pilot test, we designed and tested a new CB scale loosely modeled after the TSS. The "NewCB" measure altered the wording of this scale so that all the items gave more certainty in why a participant was choosing something as a confirmation, and created four possible answers, with two clearly offering a confirmation option, and two clearly offering a disconfirming option.

Each NewCB item offered a brief scenario followed by four response options, two of which indicate confirming responses (coded −1), and two indicating disconfirming responses (coded +1), so possible responses ranged between −2 and +2, with lower scores designating higher levels of CB. Of the six NewCB items, three were included at the pretest and three at posttest 1, with the three pretest items used again at posttest 2, and the three posttest 1 items again at the 8-week follow up posttest. The three items, which were summed to create one NewCB score, showed good internal consistency within each of the test periods (see Table 1).[5]

A second measure of CB based on Wason's (1968) Confirmation Bias Application Measure (CBAM) was also used. Participants were instructed to select information to verify clues in each item, again, using the fewest possible responses. Although Wason used just a single (card selection task) item in his CBAM, we created eight new CBAM items (with identical logical form) to assess CB. This new 9-item CBAM scale was subdivided into 3 three-item groups. Participants responded to three CBAM items at pretest, three items at each of the two posttests, and three items in the 8-week follow-up (some items were randomly chosen to repeat in different tests).

Each CBAM item offers a brief scenario followed by four response options, two of which indicate confirming or poor logic responses (coded −1 and −2) and two of which indicate partially disconfirming or fully disconfirming responses (coded +1 or +2), so possible responses ranged between −3 and +3, with lower scores indicating higher levels of CB. Of the nine CBAM items, three were included at each test period and were summed to create one CBAM score for each test. Although the internal consistency across all three test periods was acceptable, the reliabilities within two of the three tests were unsatisfactory (below .70). The three CBAM items used in posttest2 and the 8-week posttest did demonstrate good internal consistency (above .74), as did the combined alpha across all four test periods (12-item $\alpha = .84$).

5.5.3.2. FAE. To measure participants' susceptibility to FAE (i.e., the degree to which individuals rely on dispositional vs. situational attributes), we began with the "Ron's Bad Day" scenario from Riggio and Garcia (2009), and created additional clone scenarios, which included items depicting both positive and negative events. Each vignette presented a short, distinct, narrative in which the central character either suffered negative consequences from a series of poor choices and unfortunate circumstances, or enjoyed positive outcomes from superior choices and fortuitous circumstances. The scenarios were sufficiently vague as to allow participants to build their own attributions about the causes of the results in the narrative. Following Riggio and Garcia's method, after reading the passage, participants were presented with 7 situational and 3 dispositional items and asked to indicate the degree to which each played a role in the character's outcome. The results indicate consistent responses for the FAE scenarios across the three test conditions.

[5] Due to a survey error, the NewCB scale items were not included in the first two days of data collection, therefore the total number of participants for the tests below was lower for NewCB than CBAM in the tests reported below.

The second measure of FAE mitigation, adapted from Stalder (2000), required participants to watch a short video in which two individuals were ostensibly involved in an experiment concerning academic learning. In the video, the participants were tasked to play a trivia game and randomly assigned the roles of Questioner and Contestant by coin flip. The Questioner then asked a series of questions that she had written, of which the Contestant only answered three correctly. Participants were then asked to evaluate both characters. The design of the metric is such that there is no real difference between the Questioner character and the Contestant character. Recall that they are assigned their respective role through coin flip, a point made clearly evident to the viewer. Thus, any score variance between assessments of the Questioner and the Contestant is directly attributable to the participant's propensity to commit FAE. Mean scores for ratings of each character are calculated and the absolute value of the difference between the two becomes the measure for FAE.

6. Results

6.1. Analysis overview

For all analyses reported below, to capture the results from participants' posttests after the last time they played the game in the lab, we conducted a repeated measures ANCOVA collapsing Post 1 and Post 2 into a single variable called "last post". The last post variable was the repeated factor in the analysis and it contained three time periods (pre, last post, and 8-week post). Creating the last post variable enabled direct comparisons of the repeat and no repeat play conditions and it also allowed us to more accurately detect differences between repeat players and non-repeat players by considering the repetition variable as a between-subjects factor. The analysis also included duration (30 or 60 min), training type (implicit or explicit), location (University 1 or 2), and sex as between-subjects factors.

The following covariates were included: age, extroversion, openness, agreeableness, conscientiousness, emotional stability, need for closure, horizontal individualism, vertical individualism, horizontal collectivism, confirmation proneness, personal need for structure/personal fear of invalidity, sensation seeking, handedness, logic aptitude, computer comfort, gaming experience. In each case, a fully saturated model was conducted initially, significant covariates were examined, and non-significant covariates were eliminated. The reduced model was then reparameterized and reanalyzed.

6.2. Bias familiarity and knowledge

Those in the video condition reported significantly higher levels of familiarity with the biases: $t_{CB}(559) = -2.51$, $p = .001$; $t_{FAE}(559) = -3.03$, $p = .002$ than did the average participant of all the game conditions. On the bias matching test the video outperformed all game conditions, $t(58) = 2.40$, $p = .02$. This result does not support H1 which predicted the game would outperform the video Table 2.

For CB knowledge, as measured by the application test, the above repeated measures ANCOVA was conducted using duration, repetition, training type, sex, and location with the covariates extroversion and logic aptitude included. Results showed a significant between-subjects main effect for implicit/explicit training, $F(1, 527) = 8.74$, $p < .003$, $\eta^2 = .01$, and a significant quadratic within-subjects time by implicit/explicit training interaction effect, $F(1, 527) = 6.66$, $p = .01$, $\eta^2 = .01$. The time by implicit/explicit training interaction shown in Fig. 3 illustrates how the video group had greater bias knowledge at the immediate posttest, but this was

Table 2
Overall bias familiarity and matching test means and standard deviations.

	Across conditions			Explicit			Implicit			Video		
	N	*M*	SD	*n*	*M*	SD	*n*	*M*	SD	*n*	*M*	SD
Familiarity Pretest	701	1.86	1.73	373	1.95	1.75	276	1.89	1.73	52	1.12	1.41
Familiarity Postest 1	700	3.17	2.07	372	3.93[a]	1.73	276	1.76[a]	1.71	52	5.26[a]	1.32
Familiarity Postest 2	372	3.01	2.39	155	3.85[b]	2.22	165	1.50[b]	1.80			
Familiarity 8-week Postest	565	2.63	1.72	285	3.01[c]	1.62	236	2.03[d]	1.67	44	3.42[c,d]	1.61
Matching Pretest	702	2.02	1.40	374	2.04	1.43	276	1.98	1.39	52	2.13	1.25
Matching Postest 1	702	2.46	1.57	374	2.68[e]	1.57	276	1.98[e]	1.46	52	3.35[e]	1.33
Matching Postest 2	702	0.95	1.50	374	1.03	1.58	276	1.03	1.48			
Matching 8-week Postest	702	1.63	1.53	374	1.64[f]	1.60	276	1.53[g]	1.43	52	2.10[f,g]	1.56

Note: Means sharing a superscript are significantly different from one another ($p < .05$).

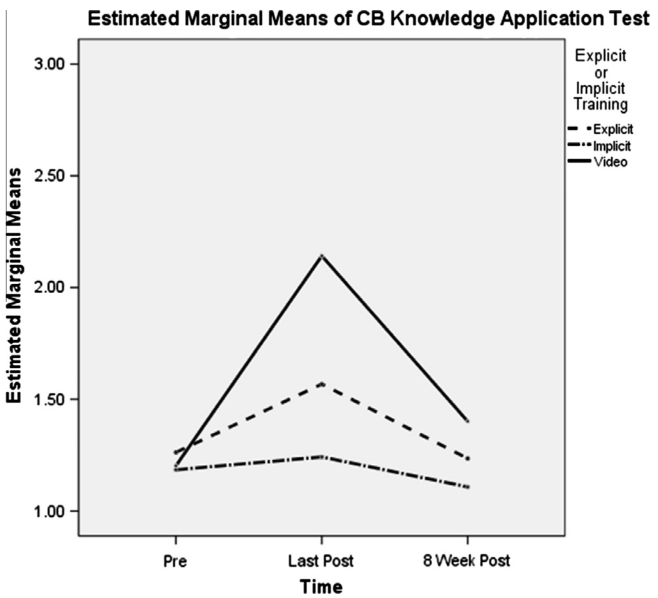

Fig. 3. Time by explicit/implicit training interaction for bias knowledge test. Higher scores indicate greater knowledge.

offset by those in the video group showing the greatest knowledge loss at 8 weeks. Comparisons between the video condition and the explicit training conditions indicated no significant differences at the 8-week posttest, $t(327) = -1.29$, $p = .20$; however, the video group was significantly different from the implicit group, $t(279) = -2.22$, $p = .03$ (see *M*s and *SD*s in Table 3). These results do not support H1, however they do provide evidence that the training gains made within the explicit training condition were longer lasting than those made by the video. Moreover, in support of H2a, comparison between explicit and implicit game conditions found the explicit game to be superior to the implicit game, $t(645) = 5.01$, $p < .001$. Concerning H3, none of the comparisons between the different duration or repetition conditions was significant.

For FAE knowledge, the model included duration, repetition, training type, sex, and location, but no covariates. Results indicated a significant quadratic within-subjects main effect for time, $F(1, 527) = 13.12$, $p < .001$, $\eta^2 = .02$, a between-subjects main effect for training type, $F(1, 527) = 7.79$, $p = .005$, $\eta^2 = .02$, and location, $F(1, 527) = 6.50$, $p = .01$, $\eta^2 = .01$, and a significant quadratic within-subjects time by duration by sex interaction, $F(1, 527) = 8.83$, $p = .003$, $\eta^2 = .02$. The main effect for time indicates FAE knowledge increased from pretest to last posttest, then decreased again at 8-week posttest. However, FAE knowledge did remain moderately higher ($p = .06$) than at the pretest. This result shows a training effect for all conditions but does not support any of the specific hypotheses.

The implicit/explicit training type main effect indicates the explicit training group had greater FAE knowledge than the implicit training group, and those in the video condition gained more knowledge than those in both the explicit and implicit training conditions (all $p < .05$). See Table 3 for Means and SDs. The location main effect indicates University 1 participants scored better overall on the FAE Knowledge test than University 2 at last post, $t(696) = 3.26$, $p = .001$ (U1: $M = 1.51$, $SD = 1.12$; U2: $M = 1.25$, $SD = .99$) and at the 8-week follow-up, $t(563) = 2.98$, $p = .003$ (U1: $M = 1.42$, $SD = 1.05$; U2: $M = 1.16$, $SD = 1.00$).

For the time by duration by sex interaction, examination of the graphs in Fig. 4 suggest both males and females performed better in the video than the game conditions. However, the video group had greater knowledge loss at 8-weeks compared to the game conditions, although this also differed by sex. For males, the 30 min condition showed knowledge loss (going from $M = 1.26$ [$SD = 1.04$] at last post to $M = 1.02$ [$SD = .90$] at 8-week follow-up) while the 60 min condition at the 8 week post actually showed improved scores over time, $t(258) = -2.65$, $p = .001$ (going from

Table 3
Bias knowledge test means and standard deviations.

	Overall			Explicit			Implicit			Video		
	N	*M*	SD	*n*	*M*	SD	*n*	*M*	SD	*n*	*M*	SD
FAE Pretest	699	1.20	0.88	372	1.24	0.89	276	1.16	0.87	51	1.08	0.87
FAE Posttest 1	699	1.36	1.05	371	1.44[a]	1.06	276	1.09[a]	0.98	52	2.15[a]	0.92
FAE Posttest 2	372	1.31	1.12	155	1.32[b]	1.16	165	1.02[b]	1.01			
FAE 8-week Postest	565	1.28	1.03	285	1.32	1.04	236	1.15[c]	1.01	44	1.70[c]	0.95
CB Pretest	699	1.20	0.82	372	1.23	0.81	276	1.19	0.81	51	1.10	0.88
CB Posttest 1	700	1.55	0.98	372	1.70[d]	0.98	276	1.26[d]	0.89	52	2.08[d]	0.90
CB Posttest 2	372	1.31	1.06	155	1.39[e]	1.08	165	0.99[e]	0.96			
CB 8-week Postest	565	1.18	0.88	285	1.22	0.89	236	1.09	0.87	44	1.41	0.87

Note: Means sharing a superscript are significantly different from one another ($p < .05$).

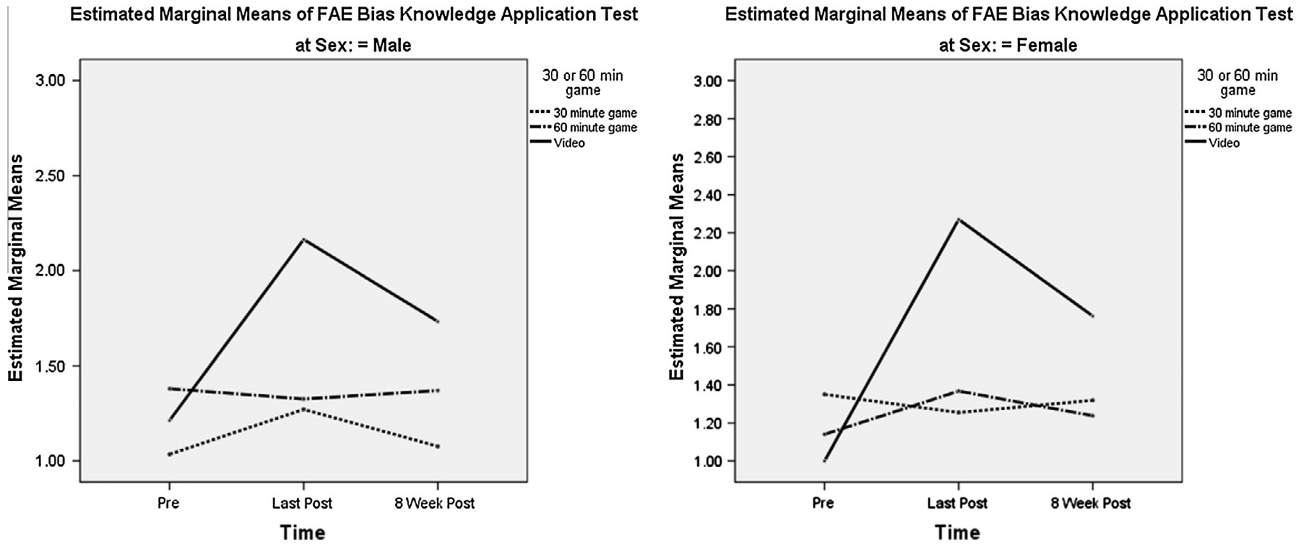

Fig. 4. Time by duration by sex interaction for FAE bias knowledge. Higher scores indicate greater knowledge.

$M = 1.26$, $SD = 1.02$, at last post to $M = 1.36$, $SD = 1.03$, at 8-week follow-up). For females this difference was not significant, $t(259) = .81$, $p = .42$. These results do not support H1 because the video outperformed the game on FAE knowledge, however, they do support H2, as the explicit training was superior to the implicit training both at the last posttest, $t(694) = 3.39$, $p = .001$ and the 8-week follow-up test $t(520) = 1.97$, $p = .05$. Finally, none of the comparisons between the different duration or repetition conditions was significant, thus H3 did not receive support.

6.3. Mitigation of CB

To test CB mitigation, we conducted two repeated-measures ANCOVAs. The first using the NewCB score (with agreeableness and need for closure retained as covariates), and the second using the CBAM score (with no covariates) as the dependent variables. Both analyses included duration, implicit/explicit training type, and repetition as between-subjects factors. There were no effects for male/female differences so sex was not included in the analysis.

The NewCB analysis yielded a significant interaction between test period and implicit/explicit training type, $F(2, 804) = 3.32$, $p = .036$, $\eta_p^2 = .01$ (see Fig. 5). In support of H2, participants in the explicit training condition showed a greater reduction in CB relative to those in the implicit training condition. Further, a *t-test* comparing the differences between explicit and implicit groups was significant at both time points: last post, $t(589) = 4.21$, $p < .001$; 8-week: $t(538) = 2.27$, $p = .02$. This effect appears to be fairly robust, as little decline between the last posttest and the 8-week posttest is evident in Fig. 5 (*M*s and *SD*s reported by training condition in Table 4).

Concerning H1, there was no overall difference between the game and the video either at the last posttest, $t(628) = -.95$, $p = .34$, or the 8-week follow-up test, $t(583) = .40$, $p = .69$. Further, the analysis revealed that none of the comparisons between the different duration or repetition conditions was significant. Thus, H2 was supported with the NewCB mitigation test for CB but H1 and H3 were not supported.

The CBAM results indicated no significant differences for any of the between-subjects factors. Moreover, there were no significant

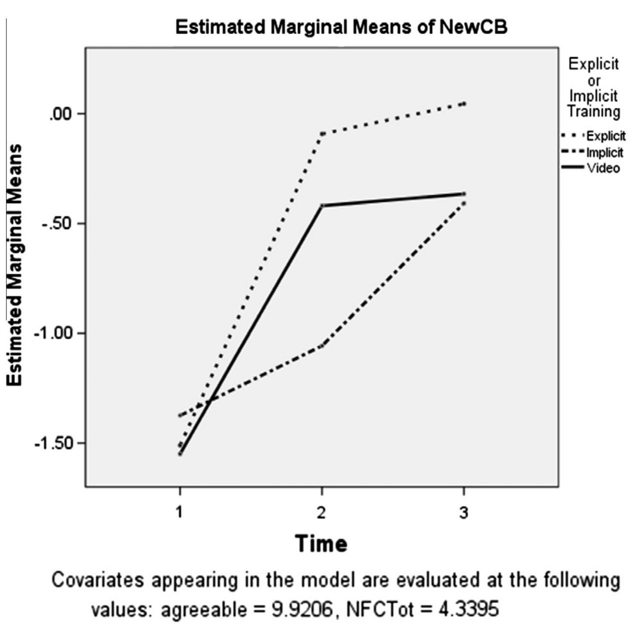

Fig. 5. Time period by training type interaction. Higher scores indicate greater mitigation.

Table 4
Means and standard deviations for NewCB by implicit/explicit training type.

	N	*M*	SD
Overall			
Pretest	489	−1.56	2.85
Last Posttest	630	−0.70	3.24
8-week Posttest	585	−0.14	3.00
Explicit			
Pretest	274	−1.58	2.95
Last Posttest	346	$-0.24^{a,b}$	3.34
8-week Posttest	297	0.12^{c}	3.05
Implicit			
Pretest	176	−1.51	2.76
Last Posttest	245	$-1.38^{a,b}$	2.98
8-week Posttest	243	-0.47^{c}	2.88
Video			
Pretest	39	−1.62	2.60
Last Posttest	39	-0.49^{b}	3.31
8-week Posttest	45	−0.04	3.13

Note: Means sharing a superscript are significantly different from one another ($p < .05$).

effects for test period, and none of the covariates significantly influenced the CBAM score. These results support none of the hypotheses.

6.4. Mitigation of fundamental attribution error (FAE)

To analyze the data, mean scores were calculated and a separate repeated measures analysis was conducted for each group of cues. Scores for dispositional attributions were calculated such that lower scores indicate less reliance on dispositional cues, whereas scores for situational attributions are calculated such that higher scores indicate increased reliance on situational cues. Tests were performed using game duration, implicit/explicit training type, and repetition, with no significant covariates being retained.

Results revealed a significant multivariate effect for time, Wilks' $\lambda = .985$, $F(2, 574) = 4.42$, $p = .012$, $\eta^2 = .015$, and a significant multivariate interaction for time by duration by repetition, Wilks' $\lambda = .989$, $F(2, 574) = 3.08$, $p = .074$, $\eta^2 = .011$. However, post hoc analysis of the effects suggests the game was not significantly different from the video (H1) and there was no significant difference between implicit and explicit conditions (H2). In comparing the many repetition and duration conditions against one another at both the last and 8-week posttests, only one significant difference emerged: The video was significantly better than the take-home condition at the 8-week posttest, $t(84)=-2.06$, $p =.04$. None of the hypotheses are supported by this analysis.

Complementary analysis for preference of situational cue scores at pretest, last posttest, and 8-week posttest examined with MANOVAs revealed a significant time × duration interaction, Wilks' $\lambda = .98$, $F(4, 1146) = 2.57$, $p = .04$, $\eta^2 = .009$, as well as a significant four-way interaction between time, implicit/explicit training, duration, and repetition, Wilks' $\lambda = .99$, $F(2, 573) = 3.65$, $p = .03$, $\eta^2 = .01$. However, none of the simple effects tests were significant, which suggests the results were largely a function of the time variable, since all the training groups increased their reliance on situational cues over time. Thus, as with the FAE knowledge test, the results suggest that there is an improvement over time for all of the conditions but none of the specific hypotheses relating to FAE were supported.

To examine how people within conditions reacted to the questioner and contestant featured in the Stalder video task (Stalder, 2000), a repeated measures MANCOVA was performed using game duration, training type, repetition, and sex. Results revealed a significant multivariate effect for time, Wilks' $\lambda = .97$, $F(1, 544) = 15.45$, $p < .001$, $\eta^2 = .03$, indicating participants in all conditions reduced their bias between the pretest and last posttest. Multivariate effects were non-significant for any of the main independent variables or covariates. Although the between-subjects effect for time shows participants in all conditions reduced their bias from last posttest ($M = 2.12$, $SD = 1.38$) to 8-week posttest ($M = 1.22$, $SD = 1.11$), $F(1, 544) = 130.51$, $p < .000$, $\eta^2 = .24$, none of the hypotheses were supported.

7. Discussion

Overall, our goal in this investigation was to demonstrate the effectiveness of game-based learning when compared to a traditional pedagogy, especially for a task as complex as mitigating the natural tendency toward cognitive bias. In addition, we investigated whether explicit or implicit instruction would be superior in teaching lessons on this type of a task, and whether the repetition of the instruction (both through repeated play sessions and longer play times) would be the most effective. The results showed promise for the use of instructional training games in this environment and provide guidance for future game developers regarding the explicitness of the instruction they need to provide.

The theoretical model that guided this study, the HSM, suggests that having greater time and the ability to think systematically about critical decisions will reduce the reliance on heuristics, and this is one of the greatest benefits that games can offer over traditional pedagogy. Games can be played at the players' own pace and the players can proceed through the levels of the game as their ability to process information allows. Unlike the speaker in the training video who keeps talking, game players can re-read information as needed and take the time to carefully consider their decisions. The fact that the increased time pressure provided in the 30-min game had little effect compared to the 60-min duration suggests players felt they had enough time to make decisions with fewer heuristics than those in the video. In contrast, the repeatability of the game made them more confident in their decisions, more familiar with the terminology being used, and gave them greater chance to practice their mitigation skills.

7.1. Training type effects

The *MACBETH* game proved to be superior to the instructional video both in attaining greater familiarity with the biases, and mitigating CB as measured by our "NewCB" test. However, the type of game proved to be quite important: The game with explicit training, when played repeatedly, was the most likely to outperform the video. Results show the explicit training version of the game to be the optimal condition for improving familiarity with the biases and knowledge of them, especially with regard to CB. The results also indicate the explicit game conditions produced greater CB mitigation than the implicit game conditions. However, this effect was qualified by a marginally significant 3-way interaction among training type, duration, and location, indicating participants at University 1 appear to have responded more favorably to explicit training in the 30-min game, whereas participants at University 2 appear to have responded more favorably to explicit training in the 60-min game. More replications are needed before drawing conclusions about the effects of the game duration on the explicitness of the training. Explicit training also produced greater FAE bias reduction and had a significant interaction with test period, whereby both explicit and implicit versions of the game successfully reduced the use of dispositional cues from pretest to the first posttest. Although allowing autonomy of learning for the user is one highly touted advantage in video games (Ryan, Rigby, & Przybylski, 2006), in this case, a more directed explicit form of training was generally more effective than an implicit form, likely due to the time constraints placed on players and the complexity of the bias training task.

7.2. Duration and repetition effects

Only partial support was found for the effect of duration on biased judgments in the pre- vs. posttest comparisons, but repetition of play had a greater role in the results. Specifically, although there was no significant effect of duration for the two FAE judgment tests, nor in the CBAM test, there was a significant main effect for NewCB on the mitigation test. This effect, however, was qualified by the three-way interaction between training type, duration, and location mentioned above.

The effects of repeat play on self-report measures of bias familiarity, knowledge, and judgments is nuanced, but when compared to single session play, repeated play offers several benefits. First, the results suggest that players in the repeat play condition feel more familiar with the biases than players in the single play condition. Moreover, this effect was qualified by a repetition by duration interaction, indicating that after 60-min, repeat players were significantly more familiar with the biases than players in all other conditions. Furthermore, the duration by training type interaction suggests explicit training combined with repeat play significantly increases players' perceptions of bias familiarity, although, players' familiarity did not translate to higher performance on the bias knowledge tests.

Repeat players exhibited more knowledge about biases than players in the single condition. However, on the FAE and CB knowledge tests, repeat players did no better (and actually worse on the matching test relative to the single-play condition). The mitigation analyses also suggest that repeat players performed better relative to single session players. Based on findings from familiarity and judgment tests, repeat play was clearly superior, providing some support for H3, even if the duration of the game by itself made little difference.

7.3. Limitations

Our comparison to the instructional video was limited by the fact that the video was created by our funding agency without our input, whereas the game was created by our team with our particular theoretical reasoning behind many of the design decisions. The game was an intelligence-based risk mitigation game, whereas the video was a series of vignettes depicting the everyday life of college students, undermining parallelism. The training video represents the current state-of-the-art method of training as provided by the funder, and follows design principles assuring "representational fidelity" (Hevner, March, Park, & Ram, 2004) and enhanced ecological validity. Although not all constraints could be controlled by the research team, future studies should investigate multiple methods with parallel content to see if the content difference or the medium is the primary driver behind these findings.

Although a training film is designed to deliver a more concentrated dose of training in a single setting, with the expectation of temporary knowledge retention at best, a video game is designed to deliver a more gradual exposure to training materials over a longer, more self-motivated, repeated series of exposures. This form of self-induced repetition should lead to greater levels of internalization and longer knowledge retention. Unfortunately, game players in the present experiment were afforded only a limited opportunity to engage in such repeated play.

Other studies have demonstrated the *MACBETH* game to be more immersive and engaging than the video (Jensen et al., submitted for publication), thus, it is reasonable to assume a more natural self-motivated, repeated exposure to game play would allow *MACBETH* and other training games like it to reach their greatest potential. It should not be surprising that a single or limited exposure to a training film or video game should have similar (i.e., not significantly different) effects. Central to the concept of gaming is the notion of self-motivated, repeated exposure, and a similar level of engagement that should not be expected from the simple viewing of a training film. Future studies should take this need for more self-instigated, repeated exposure into consideration, and also endeavor to compare like contexts so that the effects of the game itself rather than the context of game play can be isolated.

8. Conclusion

Can videogames be used as a training tool for complex problems like the mitigation of cognitive bias? This research suggests they can, however, as the results of this experiment indicate, various gaming conditions can produce a range of varied training effects, suggesting the design of future educational games needs to be tightly coupled to specific training goals focused within each game.

The present experiment compared a training video game to an instructional film, and found the game performed as well as the film in most cases, and was able to out-perform the film when explicit training and repeated play sessions were used. This was especially true for the gamers' familiarity with the two biases, and their mitigation of CB. It is important to note that MACBETH players showed greater mitigation of confirmation bias relative to those who watched the training film. Moreover, the complexity of the gaming tasks resulted in greater effectiveness for those who played the explicit training version of the game relative to those who played the implicit version.

We hope this research encourages others to explore the use of game-based learning in a range of other contexts, particularly those involving lessons or programs that are difficult-to-train, and those that offer special challenges, such as the mitigation of other forms of cognitive biases.

Acknowledgements

A previous version of this paper was presented in 2013 at the Games, Learning, Society 9.0 Conference in Madison WI. The authors thank the graduate students in the CMI program at the University of Arizona and the students of Drs. Dunbar, Jensen, Bessarabova and Miller at the University of Oklahoma for their suggestions and assistance through the development of this project. The authors would also like to thank the talented game development team at the K20 Center at the University of Oklahoma who developed the *MACBETH* game for this project.

This research was supported by the Intelligence Advanced Research Projects Activity (IARPA) via the Air Force Research Laboratory (AFRL). 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 (Contract no. FA8650-11-C-7178).

References

Abelson, R. P., & Levi, A. (1985). Decision making and decision theory. *The Handbook of Social Psychology, 1*(3), 231–309.

Chaiken, S. (1980). Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of Personality and Social Psychology, 39*(5), 752–766. http://dx.doi.org/10.1037/0022-3514.39.5.752.

Chen, S., & Chaiken, S. (1999). Dual-process theories in social psychology. In S. Chaiken & Y. Trope (Eds.), *Dual-Process Theories in Social Psychology* (pp. 73–96). New York: Guilford Press.

Ciavarro, C., Dobson, M., & Goodman, D. (2008). Implicit learning as a design strategy for learning games: Alert Hockey. *Computers in Human Behavior, 24*(6), 2862–2872.

Day, E. A., Arthur, W., & Gettman, D. (2001). Knowledge structures and the acquisition of a complex skill. *Journal of Applied Psychology, 86*(5), 1022.

Dunbar, N. E., Wilson, S., Adame, B., Elizondo, J., Jensen, M. L., Miller, C., et al. (2013). The development of a training game for the mitigation of cognitive bias: The case study of MACBETH. *International Journal of Game-Based Learning, 3*(4), 7–26. http://dx.doi.org/10.4018/ijgbl.2013100102.

Englich, B., Mussweiler, T., & Strack, F. (2006). Playing dice with criminal sentences: The influence of irrelevant anchors on experts' judicial decision making. *Personality and Social Psychology Bulletin, 32*(2), 188–200.

Gee, J. P. (2011). Reflections on empirical evidence on games and learning. In S. Tobias & J. D. Fletcher (Eds.), *Computer games and instruction* (pp. 223–232). Charlotte, NC US: IAP Information Age Publishing.

Gifford, R., & Hine, D. W. (1997). 'I'm cooperative, but you're greedy': Some cognitive tendencies in a commons dilemma. *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement, 29*(4), 257–265.

Gosling, S. D., Rentfrow, P. J., & Swann, W. B. Jr., (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality, 37*(6), 504–528. http://dx.doi.org/10.1016/s0092-6566(03)00046-1.

Habgood, M., Ainsworth, S., & Benford, S. (2005). Endogenous fantasy and learning in digital games. *Simulation and Gaming, 36*(4), 483–498.

Hartmann, T., & Klimmt, C. (2006). Gender and computer games: Exploring females' dislikes. *Journal of Computer-Mediated Communication, 11*(4).

Harvey, J. H., Town, J. P., & Yarkin, K. L. (1981). How fundamental is "the fundamental attribution error"? *Journal of Personality and Social Psychology, 40*(2), 346.
Heppner, W. L., Kernis, M. H., Lakey, C. E., Campbell, W. K., Goldman, B. M., Davis, P. J., et al. (2008). Mindfulness as a means of reducing aggressive behavior: Dispositional and situational evidence. *Aggressive Behavior, 34*(5), 486–496.
Heuer, R. J. (1999). Psychology of intelligence analysis: United States Govt Printing Office.
Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly, 28*(1), 75–105. http://dx.doi.org/10.2307/25148625.
Hill, C., Memon, A., & McGeorge, P. (2008). The role of confirmation bias in suspect interviews: A systematic evaluation. *Legal and Criminological Psychology, 13*(2), 357–371. http://dx.doi.org/10.1348/135532507x238682.
Hodgins, H. S., & Knee, C. R. (2002). The integrating self and conscious experience. In E. L. Deci & R. M. Ryan (Eds.), *Handbook of self-determination research* (pp. 87–100). Rochester, NY US: University of Rochester Press.
Hoover, D., & Whitehead, C. (1975). An experiential-cognitive methodology in the first course in management: Some preliminary results. *Simulation Games and Experiential Learning in Action, 2*, 23–25.
Jarmon, L., Traphagan, T., Mayrath, M., & Trivedi, A. (2009). Virtual world teaching, experiential learning, and assessment: An interdisciplinary communication course in Second Life. *Computers and Education, 53*(1), 169–182.
Jensen, M. L., Dunbar, N. E., Piercy, C. W., Lee, Y-H., Elizondo, J., Bessarabova, E., Twyman, N. W., Burgoon, J. K., Valacich, J. S., Adame, B. J., Miller, C. H., Wilson, S. N. (2014). Engagement in a complex digital training game: A multi-method examination (submitted for publication).
Kahneman, D. (2011). *Thinking, fast and slow*. New York, NY US: Farrar, Straus and Giroux.
Kastenmüller, A., Greitemeyer, T., Jonas, E., Fischer, P., & Frey, D. (2010). Selective exposure: The impact of collectivism and individualism. *British Journal of Social Psychology, 49*(4), 745–763. http://dx.doi.org/10.1348/014466609x478988.
Kolb, D. A. (1984). *Experiential learning: Experience as the source of learning and development* (Vol. 1). Englewood Cliffs, NJ: Prentice-Hall.
Krems, J., & Zierer, C. (1994). Are experts immune to cognitive bias? Dependence of "confirmation bias" on specialist knowledge. *Zeitschrift für experimentelle und angewandte Psychologie, 41*(1), 98–115.
Kruglanski, A. W., Webster, D. M., & Klem, A. (1993). Motivated resistance and openness to persuasion in the presence or absence of prior information. *Journal of Personality and Social Psychology, 65*(5), 861.
Masson, M. E. J., Bub, D. N., & Lalonde, C. E. (2011). Video-game training and naïve reasoning about object motion. *Applied Cognitive Psychology, 25*(1), 166–173. http://dx.doi.org/10.1002/acp.1658.
Mowday, R. T. (1981). Viewing turnover from the perspective of those who remain: The relationship of job attitudes to attributions of the causes of turnover. *Journal of Applied Psychology, 66*(1), 120–123. http://dx.doi.org/10.1037/0021-9010.66.1.120.
Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology, 2*, 175–220.
O'Brien, B. (2009). Prime suspect: An examination of factors that aggravate and counteract confirmation bias in criminal investigations. *Psychology, Public Policy, and Law, 15*(4), 315–334.
O'Sullivan, M. (2003). The fundamental attribution error in detecting deception: The boy-who-cried-wolf effect. *Personality and Social Psychology Bulletin, 29*(10), 1316–1327.
Oswald, M. E., & Grosjean, S. (2004). Confirmation bias. In R. F. Pohl (Ed.), *Cognitive illusions: A handbook on fallacies and biases in thinking, judgement and memory* (pp. 79–96). New York, NY: Psychology Press.
Raab, M. (2003). Decision making in sports: Influence of complexity on implicit and explicit learning. *International Journal of Sport and Exercise Psychology, 1*(4), 406–433.
Rassin, E. (2010). Blindness to alternative scenarios in evidence evaluation. *Journal of Investigative Psychology and Offender Profiling, 7*(2), 153–163. http://dx.doi.org/10.1002/jip.116.
Riggio, H. R., & Garcia, A. L. (2009). The power of situations: Jonestown and the fundamental attribution error. *Teaching of Psychology, 36*(2), 108–112. http://dx.doi.org/10.1080/00986280902739636.
Ryan, R. M., Rigby, C. S., & Przybylski, A. (2006). The motivational pull of video games: A self-determination theory approach. *Motivation and Emotion, 30*(4), 344–360.
Squire, K. (2003). Video games in education. *International Journal of Intelligent Games and Simulation, 2*(1), 49–62.
Stalder, D. R. (2000). Does logic moderate the fundamental attribution error? *Psychological Reports, 86*(3), 879–882. http://dx.doi.org/10.2466/pr0.2000.86.3.879.
Stalder, D. R. (2009). Competing roles for the subfactors of need for closure in committing the fundamental attribution error. *Personality and Individual Differences, 47*(7), 701–705. http://dx.doi.org/10.1016/j.paid.2009.06.005.
Todorov, A., Chaiken, S., & Henderson, M. D. (2002). The heuristic-systematic model of social information processing. In *The persuasion handbook: Developments in theory and practice* (pp. 195–211).
Tüzün, H., Yılmaz-Soylu, M., Karakuş, T., İnal, Y., & Kızılkaya, G. (2009). The effects of computer games on primary school students' achievement and motivation in geography learning. *Computers and Education, 52*(1), 68–77.
Wason, P. C. (1968). Reasoning about a rule. *Quarterly Journal of Experimental Psychology, 20*(3), 273–281. http://dx.doi.org/10.1080/14640746808400161.
Watson, G. S. (1960). More significance tests on the sphere. *Biometrika, 47*(1/2), 87–91.