Mitigation of Cognitive Bias with a Serious Game:
Two Experiments Testing Feedback Timing and Source

Norah E. Dunbar, University of California Santa Barbara, Department of Communication, Santa Barbara, California

Matthew L. Jensen, University of Oklahoma, Price College of Business and Center for Applied Social Research, Norman, Oklahoma

Claude H. Miller, University of Oklahoma, Department of Communication, Norman, Oklahoma

Elena Bessarabova, University of Oklahoma, Department of Communication, Norman, Oklahoma

Yu-Hao Lee, University of Florida, Department of Telecommunication, Gainesville, Florida

Scott N. Wilson, University of Oklahoma, K20 Center, Norman, Oklahoma

Javier Elizondo, University of Oklahoma, K20 Center, Norman, Oklahoma

Bradley J. Adame, Arizona State University, Hugh Downs School of Communication, Tempe, Arizona

Joseph Valacich, University of Arizona, Eller College of Management, Tucson, Arizona

Sara Straub, Independent Researcher, Norman, Oklahoma

Judee K. Burgoon, University of Arizona, Eller College of Management, Tucson, Arizona

Brianna Lane, Christopher Newport University, Department of Communication, Newport News, Virginia

Cameron W. Piercy, University of Central Missouri, Department of Management, Warrensburg, Missouri

David Wilson, University of Nebraska-Omaha, Department of Information Systems and Quantitative Analysis, Omaha, Nebraska

Shawn King, Independent Researcher, Norman, Oklahoma

Cindy Vincent, Salem State University, Department of Communications, Salem, Massachusetts

Ryan M. Schuetzler, University of Arizona, Eller College of Management, Tucson, Arizona

ABSTRACT

One of the benefits of using digital games for education is that games can provide feedback for learners to assess their situation and correct their mistakes. We conducted two studies to examine the effectiveness of different feedback design (timing, duration, repeats, and feedback source) in a serious game designed to teach learners about cognitive biases. We also compared the digital game-based learning condition to a professional training video. Overall, the digital game was significantly more effective than the video condition. Longer durations and repeats improve the effects on bias-mitigation. Surprisingly, there was no significant difference between just-in-time feedback and delayed feedback, and computer-generated feedback was more effective than feedback from other players.

KEYWORDS

INTRODUCTION

Proponents of digital game-based learning maintain that games and simulations can facilitate learning because they (a) cater to the digital generation of learners (Prensky, 2005), (b) allow for immersive, active learning increasing engagement and retention, and (c) encourage new forms of knowledge interaction unavailable in a traditional curricula (e.g., perspective-taking, slowing down or speeding up time processes, accessing hazardous or distant environments (Jackson, 2008). Importantly, digital games allow for immediate feedback that can help learners correct their mistakes and reward learners for making correct decisions.

The provision of feedback generally improves learning, however there are important caveats regarding how and when feedback is given. Digital games can provide feedback based on learners' pace and decision making (Azevedo & Bernard, 1995). Recent studies have examined the costs and benefits of offering feedback during instruction (Hays, Kornell, & Bjork, 2010), the timing (Butler, Karpicke, & Roediger, 2007) and the source of feedback (e.g., a teacher, parent, peer, or a computer agent in the game (Goldberg & Cannon-Bowers, 2015; Hattie & Timperley, 2007). We add to this body of research by presenting two studies exploring the effects of feedback timing (immediate vs. delayed) and feedback source (computer agents vs. human partners) in a game-based learning environment designed to teach learners about the pitfalls of cognitive biases. To test these effects, we created a serious game called MACBETH *(Mitigating Analyst Cognitive Bias by Eliminating Task Heuristics)[1]*, wherein players are tasked with detecting and preventing a series of terrorist threats by gathering and assessing intelligence data (for MACBETH development see author citation). The game focuses on knowledge and mitigation of *confirmation bias* (CB) and *fundamental attribution error* (FAE). The training effectiveness of the game was compared to a traditional instructional video explaining FAE and CB, which of course excluded feedback.

Using Feedback in a Serious Game to Mitigate Cognitive Biases

Biased information processing is often caused by the over-reliance on heuristics—defined as mental shortcuts, or simple decision rules—arising from conventional beliefs. By providing swift solutions and minimizing cognitive effort, heuristics can benefit decision-making; however, they may often also lead to insufficient consideration of relevant, diagnostic information, resulting in increased use of cognitive shortcuts associated with poor decisions and biased information processing (Tversky & Kahneman, 1974). Confirmation bias harms systematic information-processing by directing attention toward evidence that confirms existing attitudes and beliefs (Lundgren & Prislin, 1998) at the expense of weighing and examining pertinent available evidence that might otherwise disconfirm erroneous assumptions. Similarly, FAE fosters a tendency to focus on internal, dispositional explanations of others' behaviors at the expense of external, situational factors (Harvey, Town, & Yarkin, 1981) likewise hindering the decision-making process.

Cognitive biases are difficult to change: They are deeply embedded within natural cognitive processes, and people rarely recognize their biased decision-making. To mitigate bias, people must first become aware of their use of heuristics (Bornstein & Emler, 2001) for which feedback can help, thereby leading to better-informed decisions. Feedback in game-based learning can be effective when it provides players objective learning goals with clear criteria for success, along with methods for improvement to attain goals (Erhel & Jamet, 2013).

Not all feedback benefits learning: Repeated negative feedback, for instance, can lead to lowered expectations, reduced effort, and a more negative self-image (Krenn, Würth, & Hergovich, 2013). Formative and corrective outcome feedback through suggestions and guidance can help modify

thinking and behavior and improve learning (Shute, 2008). Yet, performance decrements are likely to occur if too much feedback information is presented, causing overload. Thus, both timing and quantity of feedback is critical to learning and optimal performance.

Timing has also been examined to discern the advantages of immediate versus delayed feedback, and a meta-analysis has concluded delayed feedback is generally superior in laboratory studies, since students are often required to explicitly consider and respond to it, whereas immediate feedback tends to be more effective in applied studies, such as classroom settings (Kulik & Kulik, 1988; van der Kleij, Eggen, Timmers, & Veldkamp, 2012). The amount of a "delay" varies widely in the studies with feedback being provided following an assessment, at the end of a day, or up to a week after task completion (van der Kleij et al., 2012). Although offering feedback during game play can enhance its salience, allowing players to adjust their decisions, it can also be a distraction, harming enjoyment. In-game feedback can slow game play, inhibiting goal attainment, particularly when speed of play is a basis for advancement (Ryan & Pintrich, 1997). On the other hand, despite its potential for slowing play, detailed feedback early in the process can lead to faster learning (Billings, 2010; Tsai, Tsai, & Lin, 2015). Because players can use in-task "just-in-time" (JIT) feedback to improve their performance and correct mistakes, we believe it can be more effective than feedback delayed until after task completion. Thus, we hypothesize:

H1: JIT feedback is more effective at mitigating CB and FAE than delayed feedback.

Knowledge is entwined with practice, and learning via video games is no exception (Lave & Wenger, 1991). Discovering how to play a new game takes time. Novice users can be overwhelmed with game mechanics, losing focus of the training components of the game if specific guidance and initial instruction are not provided (Serge, Priest, Durlach, & Johnson, 2013). Over time, players become more comfortable with the controls and mechanics (Dickey, 2011), allowing them to focus more on learning tactics. In a previous study testing the effects of the MACBETH game using implicit or explicit instruction, repeated play and longer duration of play were more effective than shorter or non-repeated gameplay, although the explicitness of the instruction moderated those findings (author citation). In replicating the effect of repetition and duration in mitigating CB and FAE, we hypothesize:

H2: Longer exposure to MACBETH through (a) repeated or (b) longer duration of play is more effective at mitigating CB and FAE relative to shorter duration.

EXPERIMENT 1 METHOD

Participants

A total of 508 participants (57.5% females; age: $M = 21.30$, $SD = 4.94$, range: 18-55) who fit our eligibility criteria (at least 18 years old; native English speakers) were recruited from a Midwestern university ($n = 233$) and a Southwestern University ($n = 275$) in the United States. Eleven participants were dropped prior to analyses for failing to complete all the measures, for being given incorrect measures by research staff, or for quitting gameplay before their time had expired. Overall, 411 participants (81% retention) completed the 8-week follow-up survey.

Design and Procedure

A 2 (feedback: JIT vs. delayed) × 2 (repetition: one-shot vs. repeated play) × 2 (duration: 30 vs. 60 minutes) mixed-model experiment with an offset control group (who watched an instructional video provided by our funding agency) was conducted. Descriptions of the conditions are provided below. We had no input in the design of the instructional video developed by the funding agency and did not see it until MACBETH was nearly complete. Study materials and procedures were approved and

determined to pose less than minimal risk by the internal review boards (IRBs) of both the universities and the Department of Defense.

Conditions

Feedback

Participants played either a JIT or delayed feedback version of MACBETH. The JIT version has computer mentors appearing immediately during gameplay in a box at the bottom of the screen conveying feedback on the player's actions. For example, if the player based a decision on a dispositional cue, the mentor would say: "Not quite. Look for clues about the situation next time, not dispositional cues about the suspect's personality" (see Figure 1). In the delayed feedback condition, players received the same feedback but at the end of the scenario.

Duration

Players were randomly assigned to 30- or 60-minute versions of the game, and a play clock was visible on the screen. When the time expired, players were told they must submit their final hypothesis to end the game.

Repetition

Participants in the experiment were randomly assigned to two repetition conditions: either a one-shot game session in the lab, or a repeated-play session initially in the lab, followed by a return session a week later.

The data were collected in two laboratories located in separate universities, with experimenters at each location following identical procedures. Participants first completed an online questionnaire determining their eligibility (age and English requirements) and providing their demographic and personality data; then, upon arrival at the laboratory, completed pretest measures using the Qualtrics online survey tool. Participants were randomly assigned to conditions in blocks and were asked to play MACBETH once, or watch the video once, or come for two play sessions (a week apart) in the same laboratory. Those in the game condition were also randomly assigned to duration and feedback conditions, which were held constant across play sessions for repeat players. After the game or

Figure 1. Feedback
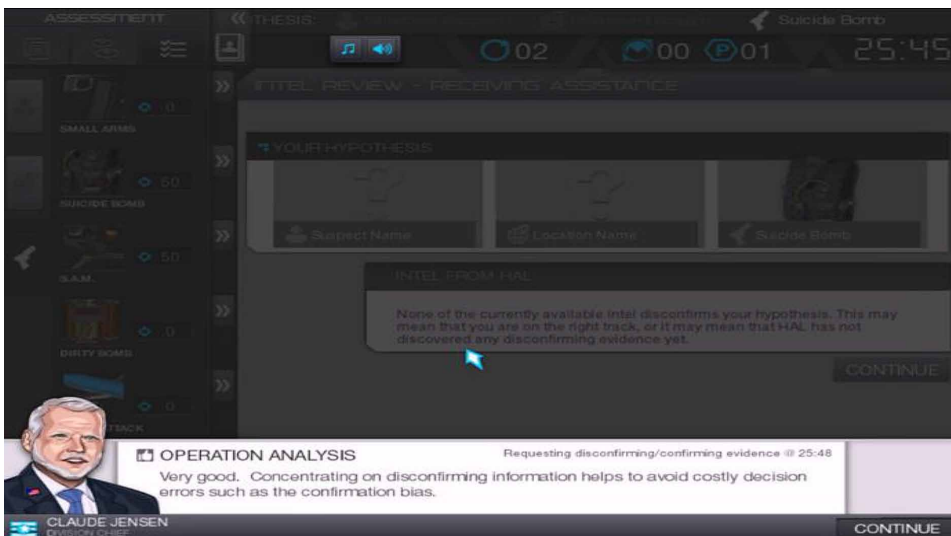
video, participants completed the post-test measures and were emailed a follow-up survey 8 weeks following their lab session. Participants were compensated $20 for each laboratory session and $30 for the follow-up survey.

Measurement

Bias Mitigation Measures

We designed and tested a new CB scale modeled after Rassin's (2010) Test Strategy Scale in which all the possible answers offered legitimate confirming and disconfirming questions that were relevant to the item's scenario. Six of these new CB measures were developed to make up two scales labeled "NewCB". Each of the two 3-item scales was used twice: They were used every other time period (pretest, posttests after both play sessions, and 8-week follow up) across the four test periods. The NewCB scale scores ranged from 0 to 28 (α = .74, .90, .92, and .90 in the four test periods).

To measure susceptibility to FAE, we began with the *Ron's Bad Day* scenario (Riggio & Garcia, (2009) and created additional scenarios to measure the degree to which individuals rely on situational vs. dispositional attributes for understanding others' behaviors. Participants saw two scenarios, one positive (e.g., Alex's successful day) and one negative (e.g., Ron's bad day). They were asked to evaluate what factors contributed to the events depicted and their scores were averaged across five dispositional items (e.g., personality, skills) and five situational items related to the scenario (e.g., the weather, contingencies). Training should result in a lower dispositional relative to situational score. The FAE scores ranged from 1 to 11 and were reliable in the four test periods for situations (α = .67, .75, .83, .91), and dispositions (α = .77, .78, .85, .92).

Experiment 1 Results

Three separate repeated-measures Analysis of variance (ANOVA) were conducted to test the hypotheses. Feedback (delayed vs. JIT), Duration (30- vs. 60-minutes), Repetition (one-shot vs. repeat-play) were entered as independent variables and the measures for CB and FAE (situation and disposition cues) were used as three separate dependent variables. To compare the video and non-repeat game condition to the repeat game condition, a "Latest Posttest" variable was created using posttest 1 for participants in the non-repeat play and video conditions, and posttest 2 for repeat play condition. Thus, the repeated measures analyses included three within-subject measures of bias mitigation: the pretest, the latest posttest, and the 8-week posttest.

Confirmation Bias Mitigation Results

The CB analysis showed that there was a significant main effect of Test Period, $F(1.90, 812.25) = 24.07$, $p < .001$, $\eta_p^2 = .05$ (Mauchly's Test of Sphericity indicated that the assumption of sphericity had been violated, therefore a Greenhouse-Geisser correction was used). The training improved the participants' CB bias mitigation ability. Pair-wise Bonferroni test showed a significant improvement from the pretest ($M = 10.22$, $SE = .24$) to the two posttest periods (latest posttest: $M = 13.01$, $SE = .35$; 8-week posttest: $M = 13.10$, $SE = 33$), but there was no significant difference between the latest posttest and the 8-week posttest.

H1 which predicted that JIT feedback improves CB mitigation ability relative to delayed feedback was not supported: The interaction between Test Period and Feedback Type was not significant, $F(1.90, 812.25) = .34$, $p = .700$, $\eta_p^2 < .01$.

H2a, predicting repeated play is more effective in mitigating CB than the single play, was supported by a significant interaction between Test Period and Repetition, $F(1.90, 812.25) = 15.41$, $p < .001$, $\eta_p^2 = .04$. Post-hoc pairwise comparison revealed the repeat play conditions (latest posttest: $M = 15.13$, $SE = .54$; 8-week posttest: $M = 14.68$, $SE = .55$) were significantly better in CB mitigation than single play (Latest posttest: $M = 11.50$, $SE = .38$; 8-week posttest: $M = 12.39$, $SE = .43$) both in the latest posttest and in the 8-week posttest (see Figure 2). In addition, both repeated and single

Figure 2. Experiment 1 results
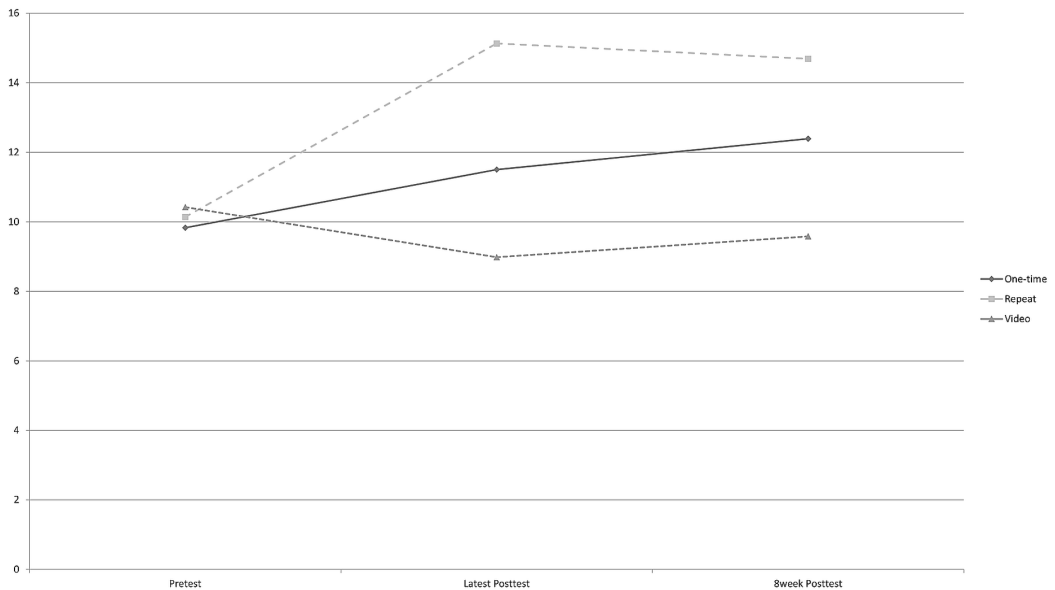
play were significantly more effective in CB mitigation than the video condition (latest posttest: M = 9.08, SE = 1.17; 8-week posttest: M = 9.84, SE = 1.11).

H2b which predicted longer duration of gameplay mitigates CB more effectively than shorter duration was not supported. The interaction between Test Period and Duration was not significant, $F(1.90, 812.25)$ = .01, p = .990, p<.001. There was no significant difference between the longer 60-minute game duration (latest posttest: M = 13.00, SE = .39; 8-week posttest: M = 13.62, SE = .44) and the shorter 30-minute game duration (latest posttest: M = 12.85, SE = .50; 8-week posttest: M = 13.47, SE = .49). However, both longer and shorter gameplay conditions were significantly more effective than the video condition (latest posttest: M = 8.98, SE = .87; 8-week posttest: M = 9.58, SE = .91) in CB mitigation.

Fundamental Attribution Error Mitigation Results

Two separate repeated-measures ANOVA were conducted to examine the hypotheses regarding FAE mitigation. The first examined whether participants decreased their reliance on dispositional cues, and the second whether participants increased their reliance on situational cues after training.

For dispositional cues, results showed a significant main effect for Test Periods, $F(1.90, 819.01)$ = 33.95, $p < .001$, η_p^2 = .07. To further investigate the effect, we performed a post-hoc Bonferroni test, which showed a significant reduction in reliance of dispositional cues between the pretest (M = 7.52, SE = .09) and the two posttest test periods (latest posttest: M = 6.64, SE = .12; 8-week posttest: M = 6.63, SE = .11). There was no significant difference between latest and 8-week posttest. Results showed across conditions participants decreased their reliance on dispositional cues after receiving the training. However, there was no significant interaction effect between Test Period and Feedback as posited by H1 ($F[1.90, 819.01]$ = 2.25, p = .911), Test Period and Repetition as posited by H2a ($F[1.90, 819.01]$ = .06, p = .940), or Test Period and Duration as posited by H2b ($F[1.90, 819.01]$ = 3.79, p =.217). These results suggest that, although playing the game decreased players' reliance of dispositional cues, neither feedback type, repetition, nor duration showed an advantage in reducing FAE.

For situational cues, there was a significant main effect for Test Periods, $F[1.68, 724.41] = 14.82$, $p < .001$, $\eta_p^2 = .03$. Post-hoc Bonferroni tests revealed no significant increase in reliance on situational cues immediately after playing the game (Pretest: $M = 7.36$, $SE = .08$, Latest Posttest: $M = 7.35$, $SE = .08$). Surprisingly, all the participants significantly decreased their reliance on situational cues after eight weeks, regardless of conditions, $M = 6.80$, $SE = .10$.

Concerning situational cues, results showed a significant interaction effect between Test Period and Feedback (H1) ($F[1.68, 724.41] = 3.64$, $p= .034$, $\eta_p^2 = .01$).

However, the difference between the two feedback conditions was not significant (JIT: $M = 70.70$, $SE = 1.03$; Delayed: $M = 71.84$, $SE= .96$), and the control video condition was significantly higher, $M = 75.08$, $SE = 2.26$. The interaction between Test Period and Repetition posited by H2a was not significant ($F[1.68, 724.41] = .59$, $p = .526$), nor was the interaction between Test Period and Duration posited by H2b ($F[1.68, 724.41] = 1.39$, $p =.250$), suggesting neither Feedback, Repetition, or Duration of game play increased reliance on situational cues.

Experiment 1 Discussion

We predicted that JIT feedback would be more effective than delayed feedback in reducing CB and FAE (reliance on dispositional cues and increase reliance on situational cues). However, we found little difference in the timing of feedback delivery, with JIT and delayed feedback performing equally well, and both outperforming the traditional instructional video in terms of reducing CB and use of dispositional cues. It could be that advantages of the faster, uninterrupted play of the delayed condition and the immediate salient feedback of the JIT conditions off-set each other. It could also be that the delay of about 20-40 minutes while the player was engaged in the scenario was not enough of a delay to make a difference. Players seemed to prefer the delayed feedback, as anecdotal comments on open-ended questions in the post survey suggested, they found the JIT feedback "annoying."

To address this issue, we modified the feedback system before Experiment 2, and tested an altered form of JIT feedback a second time. Players were given fewer positive comments from mentors and the feedback focused on corrective action to improve their performance when they made errors. The feedback quotes were also shortened wherever possible and we asked our voice actors to speed up their speech to shorten the time spent listening to feedback.

H2a posited repeated gameplay would be superior to single game play, and this was partially supported, players who played the game multiple times showed greater CB mitigation than players who played only once, but this was not true for FAE mitigation. MACBETH is a complex strategy game with a steep learning curve; players in shorter duration conditions were likely consumed with learning to navigate game mechanics, thus pointing to the efficacy of additional repeated session. Players with repeated exposure to the game were probably able to master game mechanics and better absorb the training.

H2b, which posited increased exposure to the game would enhance training, was not supported. Longer game duration provided no advantage; however, both long and short game durations were more effective than the video control condition. Comparing even the 30-minute game without repetition to the 30-minute video, the game was more effective at mitigating CB, but not FAE.

Experiment 2: Alternative Feedback Sources in the Mitigation of Cognitive Bias

Experiment 1 revealed the timing of the feedback appeared to make little difference in mitigation of CB and FAE, however, it did not address the source of the feedback, which may be a pertinent issue. Researchers have found individuals often exaggerate or understate CB when making decisions within a group (Kerschreiter, Schulz-Hardt, Mojzisch, & Frey, 2008). Tschan et al. (2009) found that having doctors display more explicit reasoning to a group when justifying their diagnosis decreased CB, and Green (1990) found that simply having to answer questions about one's decisions can eliminate CB. Even having a computer agent question one's decisions can reduce CB (Silverman, 1992). A

similar mechanism may operate for FAE, although we are unaware of any studies having tested FAE mitigation in solo vs group decision-making situations.

Michael and Chen (2006) posit that immersive collaborative virtual environments may increase students' understanding of abstract concepts. Multiplayer gaming environments encourage players to "communicate and collaborate to achieve individual and collective goals" (Dickey, 2011, p. 201), but it is unclear from the research whether multiplayer games are more conducive to learning, or whether the group distracts from an individual's learning.

We believe a multiplayer serious game can be a successful learning medium with the potential to mitigate bias more effectively than single-player training. The opportunity for players to construct their own knowledge by actively engaging with one another, beyond simply having knowledge transmitted from a screen, should lead to higher levels of learning. The success of a multiplayer version should depend on how players interact with partners, as well as the quality of feedback. Thus, we created two versions of MACBETH: One in which players traded intelligence with another player (or an artificial intelligence designed to behave like a human player, when another player was not available), and compared it to the single-player game used in Experiment 1. In addition to H3 below, we re-tested H2 to replicate the effects of repetition and duration, and again compared MACBETH to the instructional video.

H3: The Multiplayer version of MACBETH is superior to the Single Player version at mitigating CB and FAE.

Experiment 2 Method

In Experiment 2, the key variable was Player Type (single vs. multiplayer). The use of JIT feedback was held constant, and participants played either the same Single-player-JIT version of MACBETH tested in Experiment 1, or a multiplayer-JIT version, in which they played with either another human participant or a computer agent when another human player was unavailable. Experiment 2 followed the same procedures as Experiment 1 except as noted.

Participants

Participants ($N = 558$) were recruited by mass emails through the university registrar and departmental email lists, and by classroom announcements at the same two large universities. The sample of 558 participants used in the analyses included 48% females, and participants ranged from 18 to 44 years of age ($M = 21.61$, $SD = 4.89$). Of the 558 initial participants, 436 (78% retention) completed the 8-week follow-up survey. In total, 204 participants played the single-player game, and 176 played the multiplayer game, with 56 participants watching the control video. Of those who played the multiplayer version, 69 played with another human, and 107 played with a computer agent (AI), or a mix of human and computer agent.

Conditions

Player Type (Single Vs. Multiplayer)

Participants played either the same single-player-JIT version of the game used in Experiment 1, or the multiplayer-JIT version described above. Comparisons between players who played with a human or with the AI agent were not significant, therefore the two conditions were combined. Moreover, qualitative analyses of the player's comments revealed they were unaware the AI agent was not human, nor did they notice when a human player who quit was replaced by the AI.

There were several gameplay differences between the multiplayer and single player versions of the game: The single-player game in Experiment 1 had AI agents providing information, however they were not interactive, and it was clear to participants that they were not making decisions or hypotheses collaboratively. For the multiplayer version in Experiment 2, players could request assistance from

other players on even turns. The player (or agent) receiving the request had to then provide intelligence, and would receive points for doing so, as well as receive feedback from the other player they helped based on the type of information submitted.

Another difference was in the final hypothesis section. In both versions of the game, players eventually have to make a guess about the person, place, and weapon used in the terrorist attack. For the single-player version, when a player submitted a final hypothesis he or she gained points based on correct items and a bonus for the turn in which it was submitted. If the player did not have sufficient evidence to prove the hypothesis, they were penalized. For the multiplayer version, a player's final hypothesis had to be approved or rejected by the other player (or AI) they were playing with. To reject a hypothesis, a player had to submit disconfirming intelligence. If a hypothesis was approved, the submitting player received a bonus. If a hypothesis was rejected, the rejecting player received points and the submitting player received a penalty. Both players shared the final approved hypothesis, players shared points based on correct items.

Duration and Repetition

As in Experiment 1, players were randomly assigned to the 30- or 60-minute duration condition. The players were also randomly assigned to either a single play in the laboratory, repeated-play in the laboratory, or the instructional video condition.

Measures

The same bias mitigation measures in Experiment 1 were used in Experiment 2. For NewCB, Cronbach's alpha ranged from .68 to .91 for the three time periods. For dispositional FAE, alpha ranged from .85 to .93, and for situational FAE, it ranged from .77 to .88.

Experiment 2 Results

To determine the level of interdependence between human-human dyads, we conducted a series of intraclass correlations between individuals' posttest bias scores and their gaming partner's posttest bias scores as recommended by Kenny, Kashy, and Cook (2006). Results revealed no significant correlations, indicating participants' posttest bias scores were not influenced by their gaming partner's scores. Thus, players were independent of their partners, and the assumption of independence in parametric statistical tests was met.

For all analyses reported below, we conducted repeated measures ANOVA, in which Duration (30 vs. 60-min.), Repetition (one-shot vs. repeat-play) and Player Type (multiplayer vs. single-player) served as between-subject factors. To maintain comparability across conditions, the within-subjects factor (Test Period) had three levels: pretest, latest posttest (posttest 2 for the repeat players, posttest 1 for one-shot and video players), and 8-week Posttest.

Confirmation Bias Mitigation Results

To test the overall CB mitigation effect across the test periods, we conducted a single repeated-measures ANOVA. There was a significant main effect on Test Period, $F (2, 716) = 20.99, p < .001$, $\eta_p^2 = .06$. Pairwise Bonferroni comparison showed that both the latest posttest ($M = 12.04, SE = .38$) and 8-weeks posttest ($M = 12.32, SE = .37$) were higher than the pre-test score ($M = 9.70, SE = .27$), indicating, the overall trainings were effective in mitigating CB, and the mitigation effects remained even after eight weeks.

Hypothesis 3 posited playing with other players would improve the effectiveness of the trainings on bias mitigation, and the repeated-measures ANOVA results showed a significant interaction between Test Period and Player Type, $F (2, 716) = 9.48, p < .001, \eta_p^2 = .03$. However, contrary to our expectation, the single-player game ($M=12.61, SE=.47$) was significantly more effective than the multiplayer game ($M = 10.72, SE = .36$) and the video condition ($M = 8.85, SE = .73$). See Figure 3 for comparison.

Figure 3. Single-player game vs. multiplayer game
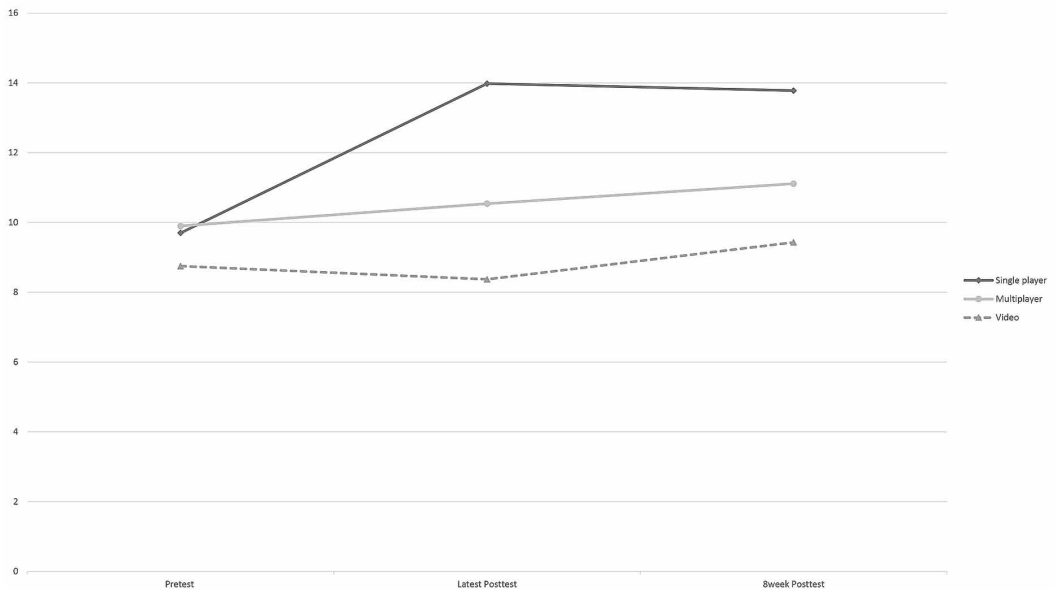
A significant Test Period × Repetition interaction, $F(2, 716) = 3.08$, $p = .046$, $\eta_p^2 = .02$, was also found, suggesting the repeat condition ($M = 12.16$, $SE = .40$) to be more effective than the single play condition ($M = 11.17$, $SE = .43$), which in turn outperformed the video condition ($M = 8.85$, $SE = .73$), with no decline at the 8-week Posttest. There was also a significant three-way Test Period × Repetition × Duration interaction, $F(2, 716) = 3.76$, $p = .024$, $\eta_p^2 = .01$, indicating the 60-minute, repeat game condition to be more effective than the other game conditions, which in turn were more effective than the video, with no drop-off from Latest Posttest to 8-week Posttest.

FAE Scenario Mitigation Results

H3 posited the multiplayer feedback design would be more effective than the single-player feedback in FAE mitigation (i.e. reduce reliance on dispositional cues and increase reliance on situational cues). However, results showed no significant difference between the single-player ($M = 6.77$, $SE = .17$) and the multiplayer ($M = 6.75$, $SE = .13$) conditions, nor the video condition ($M = 6.85$, $SE = .27$), $F(1.92, 688.28) = .21$, $p = .802$ for dispositional cues. There was no significant interaction effect between Test Period and Duration, $F(1.92, 688.28) = 2.99$, $p = .053$. There was also no significant difference in terms of Repetition (single play, repeat play, video), $F(1.92, 688.28) = 1.10$, $p = .331$.

Concerning situational cues, the goal of the study was to see if different feedback conditions would increase participants' reliance on situational cues. Omnibus results for analysis of reliance on situational cues showed a non-significant main effect for Time Period, $F(1.82, 652.77) = 1.07$, $p = .343$. No significant effects emerged for Duration (30 vs. 60-min.), $F(1.82, 652.77) = .70$, $p = .486$, Repetition (single-play, repeat-play, take-home), $F(1.82, 652.77) = .27$, $p = .745$, or Player Type (single-player vs. multiplayer), $F(1.82, 652.77) = 1.55$, $p = .215$. The data were not consistent with H3 predicting multiplayer to be more effective than single-player.

Experiment 2 Discussion

Experiment 2 replicated some of the results from Experiment 1: Longer duration and repeated play were more effective in mitigating CB than shorter duration and the single-play game, but were not more effective in mitigating FAE. The main goal of Experiment 2 was to test H3, positing the multiplayer game would outperform the single player game. However, this hypothesis was not supported. Instead,

the single-player design was more effective than the multiplayer feedback design in mitigating CB, although no differences were present for FAE. One possible explanation is that single players were more focused on their tasks. Since participants in the multiplayer version had to wait for their partners to respond to their requests, whereas participants in the single-player condition did not, it could be that less waiting time for single players led to more engagement with the training materials.

GENERAL DISCUSSION

The goal of this study was to examine how different feedback designs in a serious game may influence the effectiveness of two types of bias mitigation: confirmation bias and fundamental attribution error. We also compared the game to a professionally produced instructional video as a separate assessment of game design effectiveness. Since serious games can provide feedback to learners in a way that a static instructional video cannot, we expected the games would out-perform the video overall. This study further examined if timing and source of feedback would improve the game's effectiveness. Overall, the game performed significantly better than the video in terms of confirmation bias mitigation, with some caveats. In Experiment 1, we tested whether the timing of feedback affected bias mitigation effectiveness (H1), however results showed no significant difference between JIT and delayed feedback, although the game did reduce CB in both conditions equally well compared to the instructional video. In Experiment 2 we tested whether the feedback source affected bias mitigation effectiveness. Specifically, we hypothesized feedback from another player would be more effective than feedback from the game. Contrary to Hypothesis 3, the results showed single player feedback was significantly more effective in mitigating CB than the multiplayer feedback design; perhaps because players in the multiplayer condition had to wait for their partners to respond, thus were less engaged in the content, and may have been distracted by the interaction.

In both experiments, we tested whether longer duration of gameplay and repeated gameplay improved its effectiveness in H2a and H2b. The results were mostly consistent. The game was most effective when played for a longer duration, and with repeated play. One of the affordances of a serious game over a traditional lecture video is the former's ability to engage players for a longer duration, as well as its potential for repeat play. Through longer duration and repeated engagement, players can experiment with different solutions and observe the outcomes, thereby practicing their decision-making skills repeatedly while learning to minimize their biases.

CONCLUSION

Although feedback is generally considered beneficial for learning, our experiments tested whether the timing and the source of the feedback could affect its efficacy. Overall, the various game versions were more effective in reducing bias than the training video. The timing of the feedback—whether just in time or delayed—appeared to play little role in improving bias mitigation. However, the single-player version showed greater CB mitigation relative to the multi-player version. This suggests that in complex games with steep learning curves, like MACBETH, providing additional playing time, especially in the form of repeated learning sessions, appears to have a greater effect on learning than adding players, or adjusting the system of feedback.

REFERENCES

Azevedo, R., & Bernard, R. M. (1995). A meta-analysis of the effects of feedback in computer-based instruction. *Journal of Educational Computing Research*, *13*(2), 111–127. doi:10.2190/9LMD-3U28-3A0G-FTQT

Billings, D. R. (2010). *Adaptive feedback in simulation-based training*. Unpublished doctoral dissertation, University of Central Florida Orlando, Florida.

Bornstein, B. H., & Emler, A. C. (2001). Rationality in medical decision making: A review of the literature on doctors decision-making biases. *Journal of Evaluation in Clinical Practice*, *7*(2), 97–107. doi:10.1046/j.1365-2753.2001.00284.x PMID:11489035

Butler, A. C., Karpicke, J. D., & Roediger, H. L. (2007). The effect of type and timing of feedback on learning from multiple-choice tests. *Journal of Experimental Psychology. Applied*, *13*(4), 273–281. doi:10.1037/1076-898X.13.4.273 PMID:18194050

Dickey, M. D. (2011). World of Warcraft and the impact of game culture and play in an undergraduate game design course. *Computers & Education*, *56*(1), 200–209. doi:10.1016/j.compedu.2010.08.005

Erhel, S., & Jamet, E. (2013). Digital game-based learning: Impact of instructions and feedback on motivation and learning effectiveness. *Computers & Education*, *67*, 156–167. doi:10.1016/j.compedu.2013.02.019

Goldberg, B., & Cannon-Bowers, J. (2015). Feedback source modality effects on training outcomes in a serious game: Pedagogical agents make a difference. *Computers in Human Behavior*, *52*, 1–11. doi:10.1016/j.chb.2015.05.008

Green, D. W. (1990). Confirmation bias, problem-solving and cognitive models. *Advances in Psychology*, *68*, 553–562. doi:10.1016/S0166-4115(08)61342-4

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), 346–349. doi:10.1037/0022-3514.40.2.346

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*(1), 81–112. doi:10.3102/003465430298487

Hays, M. J., Kornell, N., & Bjork, R. A. (2010). The costs and benefits of providing feedback during learning. *Psychonomic Bulletin & Review*, *17*(6), 797–801. doi:10.3758/PBR.17.6.797 PMID:21169571

Jackson, M. (2008). Making visible: Using simulation and game environments across disciplines. In D. Davidson (Ed.), *Beyond fun: Serious games and media* (pp. 18–23). Pittsburgh, PA: ETC Press.

Kenny, D. A., Kashy, D. A., & Cook, W. L. (2006). *Dyadic data analysis*. New York: Guilford.

Kerschreiter, R., Schulz-Hardt, S., Mojzisch, A., & Frey, D. (2008). Biased Information Search in Homogeneous Groups: Confidence as a Moderator for the Effect of Anticipated Task Requirements. *Personality and Social Psychology Bulletin*, *34*(5), 679–691. doi:10.1177/0146167207313934 PMID:18310314

Krenn, B., Würth, S., & Hergovich, A. (2013). The impact of feedback on goal setting and task performance: Testing the feedback intervention theory. *Swiss Journal of Psychology/Schweizerische Zeitschrift für Psychologie/Revue Suisse de Psychologie, 72*, 79-89. doi:10.1024/1421-0185/a000101

Kulik, J. A., & Kulik, C.-L. C. (1988). Timing of feedback and verbal learning. *Review of Educational Research*, *58*(1), 79–97. doi:10.3102/00346543058001079

Lave, J., & Wenger, E. (1991). *Situated Learning: Legitimate peripheral participation*. Cambridge, UK: Cambridge University Press. doi:10.1017/CBO9780511815355

Lundgren, S. R., & Prislin, R. (1998). Motivated cognitive processing and attitude change. *Personality and Social Psychology Bulletin*, *24*(7), 715–726. doi:10.1177/0146167298247004

Michael, D. R., & Chen, S. L. (2006). *Serious games: Games that educate, train, and inform*. Boston, MA: Thomson Course Technology.

Prensky, M. (2005). Computer games and learning: Digital game-based learning. Handbook of computer game studies, 18, 97-122.

Riggio, H. R., & Garcia, A. L. (2009). The Power of Situations: Jonestown and the Fundamental Attribution Error. *Teaching of Psychology*, *36*(2), 108–112. doi:10.1080/00986280902739636

Ryan, A. M., & Pintrich, P. R. (1997). Should I ask for help?" The role of motivation and attitudes in adolescents help seeking in math class. *Journal of Educational Psychology*, *89*(2), 329–341. doi:10.1037/0022-0663.89.2.329

Serge, S. R., Priest, H. A., Durlach, P. J., & Johnson, C. I. (2013). The effects of static and adaptive performance feedback in game-based training. *Computers in Human Behavior*, *29*(3), 1150–1158. doi:10.1016/j.chb.2012.10.007

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, *78*(1), 153–189. doi:10.3102/0034654307313795

Silverman, B. G. (1992). Modeling and critiquing the confirmation bias in human reasoning. *IEEE Transactions on Systems, Man, and Cybernetics*, *22*(5), 972–982. doi:10.1109/21.179837

Tsai, F.-H., Tsai, C.-C., & Lin, K.-Y. (2015). The evaluation of different gaming modes and feedback types on game-based formative assessment in an online learning environment. *Computers & Education*, *81*, 259–269. doi:10.1016/j.compedu.2014.10.013

Tschan, F., Semmer, N. K., Gurtner, A., Bizzari, L., Spychiger, M., Breuer, M., & Marsch, S. U. (2009). Explicit Reasoning, Confirmation Bias, and Illusory Transactive Memory. *Small Group Research*, *40*(3), 271–300. doi:10.1177/1046496409332928

Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science*, *185*(4157), 1124–1131. doi:10.1126/science.185.4157.1124 PMID:17835457

van der Kleij, F. M., Eggen, T. J., Timmers, C. F., & Veldkamp, B. P. (2012). Effects of feedback in a computer-based assessment for learning. *Computers & Education*, *58*(1), 263–272. doi:10.1016/j.compedu.2011.07.020

ENDNOTES

[1]   The MACBETH game is available in the iTunes app store.

Dr. Dunbar is a Professor of Communication at the University of California Santa Barbara. She teaches courses in nonverbal and interpersonal communication, communication theory, and deception detection. She was the Principal Investigator of a $5.4 Million contract from the Intelligence Advanced Research Projects Activity in 2011-2013 and has had her research funded by the National Science Foundation, the Department of Defense, and the Center for Identification Technology Research. She has published over 50 peer-reviewed journal articles and book chapters and has presented over 90 papers at National and International conferences. Her research has appeared in top journals in her discipline including Communication Research, Communication Monographs, and Journal of Computer-Mediated Communication as well as interdisciplinary journals such as Journal of Management Information Systems and Computers in Human Behavior. She has served on the editorial board of six disciplinary journals and as the Chair of the Nonverbal Division of the National Communication Association in 2014-2016.

Matthew L. Jensen (Ph.D., University of Arizona, 2007) is an Associate Professor of Management Information Systems and a co-Director of the Center for Applied Social Research at the University of Oklahoma. Professor Jensen's interests include computer-aided decision making, knowledge management, human-computer interaction, and computer-mediated communication. He is an active member of an interdisciplinary team investigating how people attribute credibility in mediated interactions and how people filter and evaluate information they find online.

Elena Bessarabova is an Assistant Professor in the Department of Communication at the University of Oklahoma. She received her PhD in Communication from the University of Maryland, College Park in 2011. Her research foci are in the areas of social influence and cross-cultural communication. Within social influence, she is interested in resistance processes associated with psychological reactance, inoculation, and bias mitigation. She also studies the effects of emotion on resistance and attitude change. In addition, she is interested in deception as a social influence tactic, and her research on this topic focuses specifically on understanding cultural differences in deception.

Dr. Yu-Hao Lee (Ph.D., Michigan State University) is an assistant professor in the Department of Telecommunication, University of Florida. His research focuses on media psychology and information processing in interactive media environments, and online collaborations such as social media campaigns and online activism. He has received top-paper awards from the International Communication Association (ICA) and Association for Education in Journalism and Mass Communication (AEJMC). His research has been supported by external grants from the National Science Foundation (NSF), Online News Association (ONA), and AEJMC.

Scott Wilson is the Director of Innovative Technologies at the University of Oklahoma's K20 Center. This educational research center is dedicated to the research of effective teaching and learning in Kindergarten (K) to graduate (grade 20) educational environments. Dr. Wilson leads a team of researchers, designers, and developers in the application of digital game-based learning to engage learners in authentic learning experiences. He holds a PhD in Instructional Leadership and Academic Curriculum and has served as a designer, producer, and researcher of novel learning experiences and their impact on learning.

Javier Elizondo is the Production Manager for Digital Game Base Learning at the K20 Center in The University of Oklahoma. He oversees the development and implementation of educational gaming environments for educational and experimental purposes. He also manages game development of a university initiative to bring Virtual Learning Environments (VLE) to the classroom. He has extensive experience as producer of multimedia educational materials with an emphasis in educational gaming. In 2013 he produced MACBETH a game to test cognitive bias training under the sponsorship of IARPA, MACBETH went on to win Best Serious Game in bot business category as well as Special emphasis category at the Games Showcase & Challenge, at I/ITSEC 2013. Previous to his current post he served as the Communications Director for Pacific Resources for Education and Learning (PREL) and as director of JUMP Into Reading for Meaning, a Star Schools grant held by PREL that produced Cosmos Chaos! an innovative educational game to assist struggling readers with vocabulary development. Cosmos Chaos! was released on the Nintendo DS platform on the winter of 2010. Javier also served as the director of Pacific Learning Services (PLS) an instructional design company that provided multimedia educational services to a range of governments and institutions including PREL, the World Health Organization, the Asian Development Bank, the International Labor Organization, the Republic of Palau, the Federated States of Micronesia, and American Samoa. Javier Elizondo holds an MA in Communication and broadcasting by Western Illinois University and a BA in Communications Sciences by the Monterrey Institute of Technology

Bradley Adame studies how people respond to communication campaigns designed to promote disaster preparedness in areas of the country that suffer high rates of extreme weather events such as hurricanes, floods and tornados. His goal is to give guidelines for producing the most effective type of messages so that people will pay attention and take steps to protect their families and property. Prepared citizens are better able help themselves and their neighbors, in turn building stronger organizations and communities.

Brianna L. Lane (Ph.D., University of Oklahoma, 2015) is an Assistant Professor in the Department of Communication at Christopher Newport University. Her research and teaching interests include interpersonal communication within electronically mediated contexts. Her main line of research examines impression formation from online identity claims. Additionally, Brianna has conducted research regarding social support on social networking sites and the effects of cell phone use in romantic relationships.

Cameron W. Piercy, (Ph.D., University of Oklahoma, 2017) is an Assistant Professor in the Department of Management at the University of Central Missouri. He researches organizational communication, network dynamics, and computer-mediated communication. His research has been published in Computers in Human Behavior, Communication Research, and in the Proceedings of Lecture Notes in Computer Science.

David W. Wilson is an assistant professor of management information systems (MIS) at the University of Oklahoma. He received his Ph.D. in MIS from the University of Arizona. His research has appeared in MIS Quarterly, the International Journal of Human-Computer Interaction, the Journal of Nonverbal Behavior, and Communications of the AIS. He has presented his work at major IS conferences, including International Conference on Information Systems (ICIS), Americas Conference on Information Systems (AMCIS), and Hawaii International Conference on System Sciences (HICSS). His research interests include information privacy, online identity, personal disclosure, and meta-analysis methodologies.

Cindy Vincent is assistant professor in the Department of Communications at Salem State University. Her education includes a Ph.D. in Communication from the University of Oklahoma, and a M.A. in Communication and B.A. in Journalism from California State University, Sacramento. Her research primarily focuses on class representation in mainstream media and civic media creation as response.

Ryan M. Schuetzler is an Assistant Professor of Information Systems at the University of Nebraska at Omaha. He completed his Ph.D. at the University of Arizona, where he completed his dissertation on the effects of automated conversational agents on human behavior and attitudes in interviews. He earned a BS and Masters of Information Systems Management from Brigham Young University. His research interests include interpersonal deception, nonverbal behavior, intelligent agents, and human-computer interaction. His research has been published in Communications of the Association for Information Systems, Group Decision & Negotiation, and the Journal of Nonverbal Behavior, as well as numerous conferences.